

XX Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
PI WPI: 2000-013045/01.
XX DR N-PSDB: AA234464.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
XX
XX
XX Claim 1; Page 55; 67pp; English.
XX
CC The present sequence represents a human 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234464). Polymorphisms in the
CC 15 kDa selenoprotein gene are associated with susceptibility to
CC cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a 15
CC kDa selenoprotein in a subject and, if the level is below normal,
CC enhancing the level by providing additional selenium in the diet.
CC The susceptibility to cancer of a subject having an increased
CC predetermined genetic susceptibility to cancer is reduced by
CC administering the 15 kDa selenoprotein or by overexpressing the
CC selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;

Query Match 99.8%; Score 829; DB 21; Length 162;
Best Local Similarity 100.0%; Pred. No. 8.8e-88;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAGPSGGLVAFGRKLLATVLOAVSAFAGFSEACRELGFSNLLCSCDILGQPNL 60
DB 1 maagpsgglvafgrkllatvlovsaafgafseacrelgfsnllcscdillgqfnl 60
QY 61 LQLPDRCGCCOEAFETKRLVAGAILVEVCGXKLGREPOVQAVRSNDPKLRFGLQIKY 120
DB 61 lqlpdrccgcceafetkrlvagallevcgxkigrfpvqavrsdkpklrfglqiky 120
QY 121 VRGSDPVKLDDNGNIAEELSLKWNNTDVEEFLSEKLERI 162
DB 121 vrgsdpvkllddngnlaeelslkwnntdsvveeflsekleri 162

RESULT 2
ID AAY32113 standard; Protein: 162 AA.
XX
XX AAY32113;
XX
DT 01-FEB-2000 (first entry)
XX
XX Mouse 15 kDa selenoprotein.
XX
XX Selenoprotein: selenium; differential expression; tumour; mouse;
XX prostate cancer; diagnosis; polymorphism.
XX
XX Mus musculus.
XX
XX
XX Key location/Qualifiers
XX Modified-site 93
XX /label="OTHER
XX /note="selenocysteine"
XX
XX W09951637-A1.
XX
XX PD 14-Oct-1999.

XX
XX 06-APR-1999; 99MO-US07560.
XX
XX 06-APR-1998; 98US-0080850.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
XX
XX
XX WPI: 2000-013045/01.
XX DR N-PSDB: AA234469.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
XX
XX
XX Claim 1; Page 60; 67pp; English.
XX
CC The present sequence represents a mouse 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234468). Polymorphisms in the
CC human 15 kDa selenoprotein gene are associated with susceptibility
CC to cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a
CC mammalian 15 kDa selenoprotein in a subject and, if the level is
CC below normal, enhancing the level by providing additional selenium
CC in the diet. The susceptibility to cancer of a subject having an
CC increased predetermined genetic susceptibility to cancer is reduced
CC by administering a mammalian 15 kDa selenoprotein or by
CC overexpressing the selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;

Query Match 90.7%; Score 754; DB 21; Length 162;
Best Local Similarity 92.6%; Pred. No. 4.3e-79;
Matches 150; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 1 MAAGPSGGLVAFGRKLLATVLOAVSAFAGFSEACRELGFSNLLCSCDILGQPNL 60
DB 1 maagpsgglvafgrkllatvlovsaafgafseacrelgfsnllcscdillgqfnl 60
QY 61 LQLPDRCGCCOEAFETKRLVAGAILVEVCGXKLGREPOVQAVRSNDPKLRFGLQIKY 120
DB 61 lqlpdrccgcceafetkrlvagallevcgxkigrfpvqavrsdkpklrfglqiky 120
QY 121 VRGSDPVKLDDNGNIAEELSLKWNNTDVEEFLSEKLERI 162
DB 121 vrgsdpvkllddngnlaeelslkwnntdsvveeflsekleri 162

RESULT 3
ID AAM67876 standard; Protein: 92 AA.
XX
XX AAM67876;
XX
DT 25-MAR-1999 (first entry)
XX
XX Human secreted protein encoded by gene 70 clone HMSG742.
XX
XX Human: secreted protein; fusion protein; gene therapy; protein therapy;
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX developmental abnormality; foetal deficiency; blood; allergy; renal;
XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX

endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9842738-A1.

01-OCT-1998.

19-MAR-1998; 98WO-US05311.

30-MAY-1997; 97US-0050937.

21-MAR-1997; 97US-0041276.

21-MAR-1997; 97US-0041277.

21-MAR-1997; 97US-0041281.

21-MAR-1997; 97US-0042344.

30-MAY-1997; 97US-0048069.

30-MAY-1997; 97US-0048094.

30-MAY-1997; 97US-0048095.

30-MAY-1997; 97US-0048096.

30-MAY-1997; 97US-0048099.

30-MAY-1997; 97US-0048131.

30-MAY-1997; 97US-0048135.

30-MAY-1997; 97US-0048160.

30-MAY-1997; 97US-0048186.

30-MAY-1997; 97US-0048187.

30-MAY-1997; 97US-0048188.

30-MAY-1997; 97US-0048350.

30-MAY-1997; 97US-0048351.

30-MAY-1997; 97US-0048352.

30-MAY-1997; 97US-0048353.

05-AUG-1997; 97US-0054804.

(HUMA-) HUMAN GENOME SCI INC.

Brewer LA, Duan R, Ebner R, Ferrle AM, Florence KA;
Greene JM, Hu JS, Lafleur DM, Moore PA, Ni J, Olsen HS;
Rosen CA, Ruben SM, Shi Y, Young P;
WPI: 1999-070066/06.
N-PSDB; AAX00680.

New isolated human genes and the secreted polypeptides they encode -
useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders

Claim 11; Page 310; 385pp; English.

This sequence represents a secreted human protein encoded by the gene
clone detailed in the descriptor line. The gene can be used to generate
fusion proteins by linking to the gene to a human immunoglobulin Fc
portion (e.g. AAX00602) for increasing the stability of the fused
protein as compared to the human protein only.

The invention relates to 87 novel genes and their fragments (nucleic
acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
which are useful for preventing, treating or ameliorating medical
conditions e.g. by protein or gene therapy. Also, pathological
conditions can be diagnosed by determining the amount of the new
polypeptides in a sample or by determining the presence of mutations in
the new polynucleotides. Specific uses are described for each of the 87
polynucleotides, based on which tissues they are most highly expressed in
(see AAX00611 for described uses).

Sequence 92 AA;

Query Match 57.0%; Score 474; DB 20; Length 92;
Best Local Similarity 98.9%; Pred. No. 5.2e-47;
Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MAAGSGCTVPAFGKRLILATVLAQVAFGAFSSSEACRELGSSNLLCSCDLGQFNL 60
1 maagsgctvpafigrlllatvlgavsaifgaefssseacrelgfsnllcscdlgqfnl 60

61 IQLDPDRCGCCQEAQFETFKLYAGAILVEVC 92
61 Iqldpdrcgcccqeaqfekklyagaillevc 92

RESULT 4
AAV65395 standard; Protein; 95 AA.
AAV65395;
AAV65395;
01-FEB-2000 (first entry)

Human 5' EST related polypeptide SEQ ID NO:1556.

Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
gene therapy; chromosome mapping; upstream regulatory sequence;
forensic; location; development; protein synthesis; stability;
regulation; identification.

Homo sapiens.

MO9953051-A2.
21-OCT-1999.
09-APR-1999; 99WO-IB00712.
09-APR-1998; 98US-0057719.
28-APR-1998; 98US-0069047.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;
WPI: 2000-038446/03.
N-PSDB; AAZ43009.

Novel secreted protein 5' expressed sequence tag sequences used in
diagnostic, forensic, gene therapy, and chromosome mapping procedures

Claim 3; Page 818; 837pp; English.

AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
sequences, corresponding to human secreted proteins. AAV64651 to
AAV65438 represent the EST-related proteins corresponding to AAZ42265 to
AAZ43052. The 5' ESTs can be used for producing secreted human gene
products. They can be used to identify and isolate 5' untranslated
regions (UTRs) and upstream regulatory regions which control the
location, development stage, rate, and quantity of protein synthesis, as
well as stability of mRNA. The ESTs are also useful as probes for
chromosome mapping, and to obtain full length cDNA clones. The ESTs can
also be used in forensic procedures to identify individuals, or in
diagnostic procedures to identify individuals having genetic diseases
resulting from abnormal gene expression. The products may also be used in
gene therapy protocols. The nucleic acids encoding signal peptides can be
used for directing extracellular secretion of a polypeptide or the
insertion of a polypeptide into a membrane, or importing a polypeptide
into a cell. The proteins encoded by the EST sequences may be useful in
treating a variety of human conditions. Secreted proteins have
therapeutic value, and the identification of new secreted proteins is
valuable. AAZ42249 to AAZ42264 and AAV64644 to AAV64650 represent
sequences used in the exemplification of the present invention.

Sequence 95 AA;

Query Match 57.0%; Score 474; DB 21; Length 95;
Best Local Similarity 98.9%; Pred. No. 5.4e-47;
Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MAAGSGCTVPAFGKRLILATVLAQVAFGAFSSSEACRELGSSNLLCSCDLGQFNL 60
1 maagsgctvpafigrlllatvlgavsaifgaefssseacrelgfsnllcscdlgqfnl 60

AAB75288
ID AAB75288 standard; Protein; 60 AA.
XX AAB75288;
XX
XX
DT 03-APR-2001 (first entry)
XX
DE Gene 1 human secreted protein homologous amino acid sequence #107.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
XX antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnery; autoimmune disease; hyperproliferative disorder; cancer;
XX cardiovascular disorder; cerebrovascular disorder; infection;
XX nervous system disorder; ocular disorder; chemotaxis; food additive;
XX secreted protein.
XX
XX Homo sapiens.
XX
XX WO200077021-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000MO-US15135.
XX
XX 11-JUN-1999; 99US-0138632.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI: 2001-071257/08.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX Parkinson's diseases and cancers -
XX
XX Disclosure; Page 10; 530pp; English.
XX
XX This invention relates to polynucleotide sequences AAF63789 - AAF63836
XX which encode human secreted proteins AAB75260 - AAB75287. Included in the
XX invention are protein sequences AAB75288 - AAB75341 which are fragments
XX of the secreted proteins and amino acid sequences with which these
XX fragments share homology. Examples of the activities of the proteins and
XX polynucleotides and the activities of their agonists and antagonists
XX include, immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX ophthalmological; and vulnery activity. The protein and polynucleotide
XX sequences, their agonists and antagonists may be useful for treating,
XX preventing and diagnosing diseases and disorders such as autoimmune
XX diseases e.g. rheumatoid arthritis, hyperproliferative disorders
XX e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX anglogenesis, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities. Included in the invention are
XX oligonucleotides AAF63780 - AAF63788 and peptide AAB75339 which are used
XX in the identification and characterisation of the DNA and protein
XX sequences of the invention.
XX
XX Sequence 60 AA;

Query Match 36.3%; Score 302; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 AAFVRSKPKLFRLGQIKRYRGSDPVILKLDNGNIAEELSILKWNDSVEEFLSEKLERI 162
DB 1 afvrsdkpkrlfrglqikryvrgsdpvilkldngniaeelsilkwndsveeflsekleri 60
RESULT 8
ID AAB75289 standard; Protein; 60 AA.
XX AAB75289;
XX
XX AAB75289;
XX
DT 03-APR-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:108.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
XX antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnery; autoimmune disease; hyperproliferative disorder; cancer;
XX cardiovascular disorder; cerebrovascular disorder; infection;
XX nervous system disorder; ocular disorder; chemotaxis; food additive;
XX secreted protein.
XX
XX Homo sapiens.
XX
XX WO200077021-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000MO-US15135.
XX
XX 11-JUN-1999; 99US-0138632.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI: 2001-071257/08.
XX N-PSDB; AAF63789.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
XX Parkinson's diseases and cancers -
XX
XX Disclosure; Page 10; 530pp; English.
XX
XX This invention relates to polynucleotide sequences AAF63789 - AAF63836
XX which encode human secreted proteins AAB75260 - AAB75287. Included in the
XX invention are protein sequences AAB75288 - AAB75341 which are fragments
XX of the secreted proteins and amino acid sequences with which these
XX fragments share homology. Examples of the activities of the proteins and
XX polynucleotides and the activities of their agonists and antagonists
XX include, immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cytosstatic; cardiant; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX ophthalmological; and vulnery activity. The protein and polynucleotide
XX sequences, their agonists and antagonists may be useful for treating,
XX preventing and diagnosing diseases and disorders such as autoimmune
XX diseases e.g. rheumatoid arthritis, hyperproliferative disorders
XX e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX anglogenesis, nervous system disorders e.g. Alzheimer's disease,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities. Included in the invention are
XX oligonucleotides AAF63780 - AAF63788 and peptide AAB75339 which are used
XX in the identification and characterisation of the DNA and protein

CC sequences of the invention.
XX Sequence 60 AA:
SQ

Query Match 36.3%; Score 302; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 AFVRSDEKRLRGIOIKYVRSDDPKLLDNGNIAEELSLIKMNTDVEEFLSEKLERI 162
Db 1 afvrsdkpklrtgldikyvrsdpvlklddngnaeelsilkwtdsveetlseklert 60
|||||

RESULT 9
AAG08634
ID AAG08634 standard; Protein: 149 AA.
XX
AC AAG08634;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6249.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145193.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154033.
PR	20-SEP-1999;	99US-0154739.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match		
27.3%;		
Score 226.5; DB 21; Length 149;		
Best Local Similarity 32.9%;		
Pred. No. 4.6e-18;		

	Matches	48;	Conservative	37;	Mismatches	54;	Indels	7;	Gaps	4;									
Qy	17	LILATVQLQVSAFAGFEFSSEACRELFGSNNLLCSCDILLGQP-NILLQDPDCRGCCQEFA	75	:	:	:	:	:	:	:									
Db	4	lnlaastiske-----qlskcecdigifsgjalcsdchslseykkgelavsdclkcacads	59	:	:	:	:	:	:	:									
Qy	76	QEFETKKL-VAGAILEVCNGKIGRFPVOVAFTVRSDPKFLFRGLQIRYVRGSDPVLKLDDN	134	:	:	:	:	:	:	:									
Db	60	edsmkskvtysgallcwmckklvfyeylvgffieekek-fpsvkvyglfnsppllmlded	118	:	:	:	:	:	:	:									
Qy	135	GNIAEELSLTKWNTDSVEEFLSEKLE	160	:	:	:	:	:	:	:									
Db	119	gehkesiridnwkrehllygmrekyk	144	:	:	:	:	:	:	:									
RESULT 10																			
AACG08633																			
ID	AACG08633	standard; Protein; 155 AA.																	
XX																			
AC	AACG08633;																		
DT	17-OCT-2000 (first entry)																		
XX																			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 6248.																		
XX																			
KW	Protein identification: signal transduction pathway; metabolic pathway;																		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;																		
KW	termination sequence.																		
XX	Arabidopsis thaliana.																		
OS	EP1033405-A2.																		
PN																			
XX																			
PD	06-SEP-2000.																		
PF	25-FEB-2000; 2000EP-0301439.																		
XX																			
PR	25-FEB-1999;	99US-0121825.																	
PR	05-MAR-1999;	99US-0123180.																	
PR	09-MAR-1999;	99US-0123548.																	
PR	23-MAR-1999;	99US-0125788.																	
PR	25-MAR-1999;	99US-0126264.																	
PR	29-MAR-1999;	99US-0126785.																	
PR	01-APR-1999;	99US-0127462.																	
PR	06-APR-1999;	99US-0128234.																	
PR	08-APR-1999;	99US-0128714.																	
PR	16-APR-1999;	99US-0128845.																	
PR	19-APR-1999;	99US-0130077.																	
PR	21-APR-1999;	99US-0130449.																	
PR	23-APR-1999;	99US-0130510.																	
PR	23-APR-1999;	9																	

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145319.
PR 27-JUL-1999; 99US-0145318.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158655.
PR 07-OCT-1999; 99US-0158829.
PR 08-OCT-1999; 99US-0159293.
PR 12-OCT-1999; 99US-0158353.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

27.38; Score 226.5; DB 21; Length 155;

[illegible]

PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137602
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	14-JUN-1999	9905-0138847
PR	15-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139650
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139819
PR	22-JUN-1999	9905-0139867
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	30-JUN-1999	9905-0140991
PR	01-JUL-1999	9905-0141287
PR	01-JUL-1999	9905-0141442
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143542
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144086
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144632
PR	20-JUL-1999	9905-0144684
PR	21-JUL-1999	9905-0144614
PR	21-JUL-1999	9905-0144508
PR	21-JUL-1999	9905-0145085
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145182
PR	23-JUL-1999	9905-0145145
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145524
PR	26-JUL-1999	9905-0145576
PR	27-JUL-1999	9905-0145513
PR	27-JUL-1999	9905-0145518
PR	28-JUL-1999	9905-0145519
PR	28-JUL-1999	9905-0145521
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146388
PR	03-AUG-1999	9905-0146589
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143297.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.	PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.	PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.	PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.	PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.	PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.	PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.	PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.	PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.	PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.	PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.	PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.	PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149338.	PR	16-AUG-1999;	99US-0149338.
PR	17-AUG-1999;	99US-0149175.	PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.	PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.	PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.	PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.	PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.	PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.	PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.	PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.	PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151085.	PR	27-AUG-1999;	99US-0151085.
PR	27-AUG-1999;	99US-0151086.	PR	27-AUG-1999;	99US-0151086.
PR	27-AUG-1999;	99US-0151080.	PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.	PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.	PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.	PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.	PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.	PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.	PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.	PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.	PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.	PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.	PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.	PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.	PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.	PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.	PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.	PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.	PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.	PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.	PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.	PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.	PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159283.	PR	13-OCT-1999;	99US-0159283.
PR	13-OCT-1999;	99US-0159295.	PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.	PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.	PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.	PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.	PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.	PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0160741.	PR	18-OCT-1999;	99US-01

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142380.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

[illegible]

CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells; to treat disorders of haematopoietic cells, autoimmune
disorders, allergic reactions, graft versus host disease and organ
rejection, modulate haemostatic or thrombolytic activity, modulate
inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections, the peptides, nucleotides, antibodies, and
agonists and antagonists may be also be used in drug screens, AAC78449 to
CC AAC78457 and ABA444240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 34 AA:

Query Match 20.5%; Score 170; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. NO. 2.1e-12;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 129 KLDDNGNIAEELSLTKWNTDSVEEFLSEKLERI 162
| | | | | | | | | | | | | | | | | | | | | |
db 1 klddngniaeelsltkwntdsveefisekleri 34

Search completed: September 18, 2002, 16:32:01
Job time: 3876 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:32:05 ; Search time 30.41 Seconds
(without alignments)
591.712 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 162
Sequence: 1 MAAGPSGCLVPAGFKRLLLA.....ILKNTDSVEEELSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq 032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Query & Length of match SUMMARIES

Result No.	Score	Match Length	ID	Description
1	161	99.4	162	Human 15 kDa selen
2	95	58.6	162	Mouse 15 kDa selen
3	77	47.5	20	Human secreted pro
4	77	47.5	95	Human 5' EST relat
5	77	47.5	124	Human prostate can
6	60	37.0	60	Gene 1 human secre
7	60	37.0	22	Human secreted pro
8	34	21.0	34	Human cancer assoc
9	28	17.3	50	Human 5' EST relat
10	8	4.9	142	Drosophila melanog
11	8	4.9	303	Propionibacterium

12	7	4.3	46	20	AAV27650	Human secreted pro
13	7	4.3	120	21	AAU49944	Propionibacterium
14	7	4.3	149	22	AAU08634	Arabidopsis thalia
15	7	4.3	151	12	AAU13304	CFR 556 del A. H
16	7	4.3	155	21	AAU08633	Arabidopsis thalia
17	7	4.3	155	21	AAU43258	Arabidopsis thalia
18	7	4.3	156	22	AAU42036	Human polypeptide
19	7	4.3	158	21	AAU08632	Arabidopsis thalia
20	7	4.3	159	21	AAU43257	Arabidopsis thalia
21	7	4.3	159	21	AAU43256	Arabidopsis thalia
22	7	4.3	190	22	AAU07527	Novel human diagno
23	7	4.3	281	21	AAU07226	Arabidopsis thalia
24	7	4.3	319	21	AAU07225	Arabidopsis thalia
25	7	4.3	346	18	AAU23088	Pyrodicticum sp. es
26	7	4.3	364	14	AAU31014	Prod. of a probe o
27	7	4.3	424	22	AAU51207	Rat Csk binding pr
28	7	4.3	426	22	AAU67773	Drosophila melanog
29	7	4.3	464	22	AAU59876	Drosophila melanog
30	7	4.3	492	12	AAU13236	CFR Q493X. Homo
31	7	4.3	503	19	AAU53232	Cryptosporidium pa
32	7	4.3	510	19	AAU38289	Murine N-acetylgl
33	7	4.3	541	12	AAU13296	CFR G542X. Homo
34	7	4.3	660	18	AAU44120	Human ADAM 12 pred
35	7	4.3	734	22	AAU91651	C glutamicum prote
36	7	4.3	734	22	AAU76657	Corynebacterium gl
37	7	4.3	809	22	AAU09902	Novel human diagno
38	7	4.3	836	16	AAU79835	Truncated cystic f
39	7	4.3	1091	12	AAU13303	CFR Y1092X. Homo
40	7	4.3	1190	12	AAU13308	CFR 3659 del C.
41	7	4.3	1255	22	AAU84961	Shrimp white spot
42	7	4.3	1284	22	AAU62713	Drosophila melanog
43	7	4.3	1476	20	AAU33968	CFR protein seque
44	7	4.3	1479	12	AAU1602	Mutant cystic fibr
45	7	4.3	1479	12	AAU13231	CFR delta 1507.

ALIGNMENTS

RESULT 1	AAV32112	standard: Protein: 162 AA.
ID	AAV32112	
XX	AAV32112	
AC	AAV32112	
XX	01-FEB-2000	(first entry)
DT	XX	
XX	Human 15 kDa selenoprotein.	
DE	XX	
XX	Selenoprotein; selenum; differential expression; tumour; human;	
KW	prostate cancer; diagnosis; polymorphism.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
PH	Key	Location/Qualifiers
FT	Peptide	1..26
FT	Protein	/note="signal peptide"
FT	Protein	27..162
FT	Protein	/note="mature protein (specifically claimed in Claim 12)"
FT	Modified-site	93
FT	Modified-site	/label="OTHER"
FT	Modified-site	/note="selenocysteine"
PN	W09951637-A1.	
XX	14-OCT-1999.	
PD	XX	
XX	06-APR-1999;	99WO-US07560.
PF	XX	
XX	06-APR-1998;	98US-0080850.
PR	XX	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	

XX Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
PI
XX WPI: 2000-013045/01.
DR N-PSDB: AA234464.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
XX
PS Claim 1; Page 55; 67pp: English.
XX
CC The present sequence represents a human 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234464). Polymorphisms in the
CC 15 kDa selenoprotein gene are associated with susceptibility to
CC cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a 15
CC kDa selenoprotein in a subject and, if the level is below normal,
CC enhancing the level by providing additional selenium in the diet.
CC The susceptibility to cancer of a subject having an increased
CC predetermined genetic susceptibility to cancer is reduced by
CC administering the 15 kDa selenoprotein or by overexpressing the
CC selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;
XX
Query Match 99.4%; Score 161; DB 21; Length 162;
Best Local Similarity 100.0%; Pred. No. 2.9e-150;
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MAAPSGCLVPAFGKRIILAVIQAVSAFGAFSESEARELGFSSNLLCSCDLIGQNL 60
DB 1 maapsgclvpafigkriilavqlavsafigaefseacrelgfsnllcscdligqnl 60
QY 61 LQLPDPKCCQCEAEQETFKLYAGAILVEVCGXKLGRRPPVOAFAVRSRDKPLRGLQIKY 120
DB 61 lqlpdpkccqceaeqetfklyagailvevcgxlgrfpvqafvrsdkpqlrfglqiky 120
QY 121 VRGSDPYVKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
DB 121 vrgsdpyvklldngniaeelsilkwmntdsvveeflsekleri 162
XX
RESULT 2
AAV32113
ID AAV32113 standard; Protein: 162 AA.
XX
AC AAV32113;
XX
DT 01-FEB-2000 (first entry)
XX
DE Mouse 15 kDa selenoprotein.
XX
KM Selenoprotein; selenium; differential expression; tumour; mouse;
KM prostate cancer; diagnosis; polymorphism.
XX
OS Mus musculus.
XX
XX Key Location/Qualifiers
FH 93
FT Modified-site /label= OTHER
FT /note= "selenocysteine"
XX
XX MO9951637-A1.
XX
XX 14-OCT-1999.

XX
XX 06-APR-1999; 99WO-US07560.
XX
XX 06-APR-1998; 98US-0080850.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;
PI
XX WPI: 2000-013045/01.
DR N-PSDB: AA234469.
XX
PT New isolated selenoprotein polypeptides, used to develop products for
PT detecting susceptibility to or treating cancers e.g. prostate cancers
PT
XX
PS Claim 1; Page 60; 67pp: English.
XX
CC The present sequence represents a mouse 15 kDa selenoprotein that
CC is differentially expressed in cancer cells, such as cancer cells.
CC It includes a selenocysteine residue at position 93 that is encoded
CC by a TGA codon in the cDNA (see AA234468). Polymorphisms in the
CC human 15 kDa selenoprotein gene are associated with susceptibility
CC to cancer. Claimed methods of detecting the presence, or of
CC quantifying the level of expression, of the 15 kDa selenoprotein
CC involve contacting a sample with a specific binding agent, such as
CC a polyclonal antibody or monoclonal antibody, that specifically
CC binds to the 15 kDa selenoprotein. A claimed method for dietary
CC regulation comprises detecting an abnormally low expression of a
CC mammalian 15 kDa selenoprotein in a subject and, if the level is
CC below normal, enhancing the level by providing additional selenium
CC in the diet. The susceptibility to cancer of a subject having an
CC increased predetermined genetic susceptibility to cancer is reduced
CC by administering a mammalian 15 kDa selenoprotein or by
CC overexpressing the selenoprotein using gene therapy.
XX
SQ Sequence 162 AA;
XX
Query Match 58.6%; Score 95; DB 21; Length 162;
Best Local Similarity 100.0%; Pred. No. 2.4e-85;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 67 CRCCQGEAEQETFKLYAGAILVEVCGXKLGRRPPVOAFAVRSRDKPLRGLQIKVRSRDP 126
DB 67 crccqgeaeqetfklyagailvevcgxlgrfpvqafvrsdkpqlrfglqikvrsdp 126
QY 127 VKLLDNGNIAEELSLKWMNTDVEEFLSEKLERI 162
DB 127 vkllldngniaeelsilkwmntdsvveeflsekleri 162
XX
RESULT 3
AAW67876
ID AAW67876 standard; Protein: 92 AA.
XX
AC AAW67876;
XX
DT 25-MAR-1999 (first entry)
XX
DE Human secreted protein encoded by gene 70 clone HMSG42.
XX
KM Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal;
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
XX Homo sapiens.
XX
XX

PN W09842738-A1.
 XX 01-OCT-1998.
 PD
 XX
 PF 19-MAR-1998; 98MO-US05311.
 XX
 PR 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048355.
 PR 05-AUG-1997; 97US-0054804.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Duan R, Ebner R, Ferrle AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DM, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;
 XX
 XX MPI. 1999-070066/06.
 DR N-PSDB; AAX00680.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 11: Page 310; 385pp; English.
 XX
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAW67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).
 CC
 XX
 SQ Sequence 92 AA:

Query Match 47.5%; Score 77; DB 20; Length 92;
 Best Local Similarity 100.0%; Pred. No. 7.2e-68;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLILATVLAQAVSAFGAEESSACRELGFSNNLSCSDLLGQNLQLDPDRCGCCOEFA 75
 DB 16 RLILATVLAQAVSAFGAEESSACRELGFSNNLSCSDLLGQNLQLDPDRCGCCOEFA 75
 OY 76 QFETKRLVAGAILVEVCG 92
 DB 76 QFETKRLVAGAILVEVCG 92

RESULT 4
 AAY65395
 ID AAY65395 standard; Protein; 95 AA.
 AC AAY65395;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST related polypeptide SEQ ID NO:1556.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.
 XX
 OS Homo sapiens.
 XX
 PN W09953051-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 09-APR-1999; 99MO-IB00712.
 XX
 PR 09-APR-1998; 98US-0057719.
 PR 28-APR-1998; 98US-0069047.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 PI MPI. 2000-038446/03.
 DR N-PSDB; AAZ43009.
 XX
 PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
 XX
 PS Claim 3; Page 818; 837pp; English.
 XX
 CC AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
 CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 95 AA:

Query Match 47.5%; Score 77; DB 21; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.4e-68;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLILATVLAQAVSAFGAEESSACRELGFSNNLSCSDLLGQNLQLDPDRCGCCOEFA 75
 DB 19 RLILATVLAQAVSAFGAEESSACRELGFSNNLSCSDLLGQNLQLDPDRCGCCOEFA 78
 OY 76 QFETKRLVAGAILVEVCG 92

Db 79 qfckklyagailevcg 95

|||||

RESULT 5
AAB57132
ID AAB57132 standard; Protein: 124 AA.
XX
AC AAB57132;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1710.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytostatic; cardioprotective; immunomodulatory; muscular;
vulnerary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO20005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
DR N-PSDB: AAF16335.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 2183-2184; 2338pp: English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioprotective, immunomodulatory, muscular, vulnerary, gastrointestinal,
CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF15506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 124 AA;

Query Match 47.5%; Score 77; DB 21; Length 124;
Best Local Similarity 100.0%; Pred. No. 9, 3e-68;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 RLILATYQVANSARGAERSSACRELGRSSMLSSCDLLQGFNLLQDPDRCGCCQGEA 75
Db 48 lllllltvlgavsaifgaefssacrelgrfssnllcscdlilgfnllqdpdrcgscgea 107
QY 76 QEFTRKLYAGAILLEVCG 92
|||||

Db 108 qfckklyagailevcg 124

|||||

RESULT 6
AAB75288
ID AAB75288 standard; Protein: 60 AA.
XX
AC AAB75288;
XX
DT 03-APR-2001 (first entry)
XX
DE Gene 1 human secreted protein homologous amino acid sequence #107.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; nocotropic;
antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnerary; autoimmune disease; hyperproliferative disorder; cancer;
cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
secreted protein.
XX
OS Homo sapiens.
XX
PN WO200077021-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15135.
XX
PR 11-JUN-1999; 99US-0138632.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoullis GA;
XX
DR WPI: 2001-071257/08.
DR Parkinson's diseases and cancers -
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Disclosure; Page 10; 530pp: English.
XX
CC This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC invention are protein sequences AAB75288 - AAB75341 which are fragments
CC of the secreted proteins and amino acid sequences with which these
CC fragments share homology. Examples of the activities of the proteins and
CC polynucleotides and the activities of their agonists and antagonists
CC include, immunosuppressive; antiarthritic; antirheumatic;
CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
CC nocotropic; neuroprotective; antibacterial; virucide; fungicide;
CC ophthalmological; and vulnerary activity. The protein and polynucleotide
CC sequences, their agonists and antagonists may be useful for treating,
CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.
XX
SQ Sequence 60 AA;

Query Match 37.0%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.6e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 AFVRSDBPKLFRGLQIKYVSGSDPVKLLDDNGNIAEELSLKMNWDVSEEFLEKLERI 162
DB 1 afvrsdkpklfrglqikkyvsgsdpvkllddngniaeelslkwnldsvseeflsekleri 60

RESULT 7
AAB75289
ID AAB75289 standard; Protein; 60 AA.
XX
XX AAB75289;
DT 03-APR-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:108.
DE
XX Human; immunosuppressive; antiarthritic; antirheumatic; nootropic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; autoimmune disease; hyperproliferative disorder; cancer;
KW cardiovascular disorder; cerebrovascular disorder; infection;
KW nervous system disorder; ocular disorder; chemotaxis; food additive;
secreted protein.
XX
XX Homo sapiens.
OS
XX WO200077021-A1.
PM
XX 21-DEC-2000.
PD
XX 01-JUN-2000; 2000WO-US15135.
PF
XX 11-JUN-1999; 99US-0138632.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
PI
XX WPI: 2001-071257/08.
DR
XX N-PSDB: AAF63789.
DR
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
XX Disclosure: Page 10; 530pp; English.

This invention relates to polynucleotide sequences AAF63789 - AAF63836
which encode human secreted proteins AAB75280 - AAB75287. Included in the
invention are protein sequences AAB75288 - AAB75341 which are fragments
of the secreted proteins and amino acid sequences with which these
fragments share homology. Examples of the activities of the proteins and
polynucleotides and the activities of their agonists and antagonists
include, immunosuppressive; antiarthritic; antirheumatic;
antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
nootropic; neuroprotective; antibacterial; virucide; fungicide;
ophthalmological; and vulnary activity. The protein and polynucleotide
sequences, their agonists and antagonists may be useful for treating,
preventing and diagnosing diseases and disorders such as autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders
e.g. neoplasms of the breast or liver, cardiovascular disorders
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
angiogenesis, nervous system disorders e.g. Alzheimer's disease,
infections caused by bacteria, viruses and fungi and ocular disorders
e.g. corneal infection. The polypeptides can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.
XX
XX Sequence 60 AA;
SQ

Query Match 37.0%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.6e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 103 AFVRSDBPKLFRGLQIKYVSGSDPVKLLDDNGNIAEELSLKMNWDVSEEFLEKLERI 162
DB 1 afvrsdkpklfrglqikkyvsgsdpvkllddngniaeelslkwnldsvseeflsekleri 60

RESULT 8
AAB44006
ID AAB44006 standard; Protein; 34 AA.
XX
XX AAB44006;
AC
XX 08-FEB-2001 (first entry)
DT
XX Human cancer associated protein sequence SEQ ID NO:1451.
DE
XX
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.
XX
XX Homo sapiens.
OS
XX WO200055350-A1.
PM
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05882.
PF
XX 12-MAR-1999; 99US-0124270.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-587533/55.
DR
XX N-PSDB: AAC78215.
DR
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
PT
XX
XX Claim 11; Page 2132; 2352pp; English.

AAC77607 to AAC78448 encode the human cancer associated proteins given
in AAB43398 to AAB44239. The proteins can have activities based on the
tissues and cells the genes are expressed in. Example of activities
include: cytostatic; proliferative; vulnary; immunomodulator;
antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
the present invention may be used to treat immune disorders by activating

CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 34 AA;

Query Match 21.0%; Score 34; DB 21; Length 34;
 Best Local Similarity 100.0%; Pred. No. 5.7e-26;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 129 KLIDNGNIABEILSLKWNDSVEEPLSEKLERI 162
 ||||||||||||||||||||||||||||||||
 Db 1 KLIDNGNIAEELSLKWNDSVEEPLSEKLERI 34

RESULT 9

ID AAY65389 standard; Protein; 50 AA.

AC AAY65389;

DT 01-FEB-2000 (first entry)

DE Human 5' EST related polypeptide SEQ ID NO:1550.

KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.

OS Homo sapiens.

PN WO9953051-A2.

PD 21-OCT-1999.

PF 09-APR-1999; 99WO-IB00712.

PR 09-APR-1998; 98US-0057719.

PR 28-APR-1998; 98US-0069047.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-038446/03.

DR N-PSDB; AAZ43003.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures

PS Claim 3; Page 816-817; 837pp; English.

XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
 CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be

CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AAZ42249 to AAZ42264 and AAY6464 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 50 AA;

Query Match 17.3%; Score 28; DB 21; Length 50;
 Best Local Similarity 100.0%; Pred. No. 6.4e-20;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 KYRGSDPVLKIDNGNIAEELSLK 146
 ||||||||||||||||||||||||||||
 Db 20 KYRGSDPVLKIDNGNIAEELSLK 47

RESULT 10

ID ABB63319 standard; Protein; 142 AA.

AC ABB63319;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 16749.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL07422.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

PS interactions -
 PS Disclosure; SEQ ID NO 16749; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 142 AA;

Query Match 4.9%; Score 8; DB 22; Length 142;

Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 IQIKYRG 123
|11111111|
Db 69 lqtkyrg 76

RESULT 11

AAU46002
ID AAU46002 standard; Protein: 303 AA.

XX AAU46002;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #6898.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skelley YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX MPI: 2001-616774/71.

XX N-PSDB; AAS59529.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -

XX Example 1; SEQ ID No 7197; 1063pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 303 AA;

Query Match 4.9%; Score 8; DB 22; Length 303;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLERI 162
|11111111|
Db 167 lseklert 174

RESULT 12

AAAY27650
ID AAAY27650 standard; Protein: 46 AA.

XX AAAY27650;

XX 30-JUL-1999 (first entry)

XX Human secreted protein encoded by gene No. 84.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX developmental abnormality; foetal deficiency; blood; allergy; renal;
XX immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX WO9924836-A1.

XX 20-MAY-1999.

XX 04-NOV-1998; 98WO-US23435.

XX 17-NOV-1997; 97US-0066100.

XX 07-NOV-1997; 97US-0064900.

XX 07-NOV-1997; 97US-0064908.

XX 07-NOV-1997; 97US-0064911.

XX 07-NOV-1997; 97US-0064912.

XX 07-NOV-1997; 97US-0064983.

XX 07-NOV-1997; 97US-0064984.

XX 07-NOV-1997; 97US-0064985.

XX 07-NOV-1997; 97US-0064987.

XX 07-NOV-1997; 97US-0064988.

XX 17-NOV-1997; 97US-0066090.

XX 17-NOV-1997; 97US-0066094.

XX 17-NOV-1997; 97US-0066095.

XX 17-NOV-1997; 97US-0066089.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Edner R, Endress GA, Feng P, Janat F;
XX Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
XX Ruben SM, Shi Y, Soppet DR, Wei Y;

XX MPI: 1999-337740/28.

XX N-PSDB; AAX85016.

XX New human secreted proteins and coding sequences useful for treating
XX disorders of the immune system and hyperproliferative disorders
XX Claim 11; Page 397; 507pp; English.

XX This sequence represents a secreted human protein encoded by the gene
XX clone detailed in the descriptor line. The gene can be used to generate
XX fusion proteins by linking to the gene to a human immunoglobulin FC
XX portion (e.g. AAX84924) for increasing the stability of the fused
XX protein as compared to the human protein only.
XX The invention relates to 125 novel genes and their fragments (nucleic
XX acid sequences: AAX84933-X85057; amino acid sequences AAY7567-727933)
XX which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 125
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX04933 for described uses).

CC
XX
SQ Sequence 46 AA;

Query Match 4.3%; Score 7; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 RGSDDL 128
Db 18 rgsdpl 24

RESULT 13
AAU49944 standard; Protein: 120 AA.
XX
AC AAU49944;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #10840.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neutroprotectant.
XX
OS Propionibacterium acnes.
XX
PN W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001MO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
PI Sleiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
DR N-PSDB; AAS59546.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Example 1; SEQ ID NO 11139; 10699P; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 120 AA;

Query Match 4.3%; Score 7; DB 22; Length 120;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGPSGCL 9
Db 11 agpsgcl 17

RESULT 14
AAG08634 standard; Protein: 149 AA.
XX
AC AAG08634;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6249.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138450.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142555.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 4.3%; Score 7; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GALLEVC 91
Db 70 gallevc 76

RESULT 15

AAR13304
ID AAR13304 standard; Protein; 151 AA.

AC AAR13304;

DT 14-OCT-1991 (first entry)

DE CFTF 556 del A.

Deletion; mutant; diagnosis; antibodies; drug therapy.

OS Homo sapiens.

Key Location/Qualifiers

FT Domain

FT Domain /label= membrane-spanning_domain

FT Domain /label= membrane-spanning_domain

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

FT Modified-site /label= phosphorylation_site

FT Modified-site /note= "by protein kinases C"

Query Match 4.3%; Score 7; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 ISEKIER 161
Db 49 IseKier 55

Search completed: September 18, 2002, 16:36:56
Job time: 291 sec

Sequence 151 AA;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:35:25 ; Search time 23.9 Seconds
(without alignments)
1172.602 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 162
Sequence: 1 MAAGPSCGLVPARFKRLLLA.....ILKWNDSVEEPLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP rivirus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	146	90.1	162	4	O9GZM0	O9GZM0 homo sapien
2	95	58.6	162	11	O9ERR7	O9ERR7 mus musculu
3	78	48.1	162	11	O923V8	O923V8 rattus norv
4	77	47.5	92	4	O9BS64	O9BS64 homo sapien
5	9	5.6	152	5	O9N4C6	O9N4C6 caenorhabdi
6	8	4.9	97	16	O9A1M1	O9A1M1 streptococc
7	8	4.9	142	5	O9VVJ7	O9VVJ7 drosophila
8	8	4.9	295	16	O9S578	O9S578 pseudomonas
9	8	4.9	365	11	O9JKU7	O9JKU7 mus musculu
10	8	4.9	641	12	O9WY13	O9WY13 breviabacter
11	7	4.3	141	17	O96YR9	O96YR9 sulfolobus
12	7	4.3	158	10	O9SYL3	O9SYL3 arabidopsis
13	7	4.3	159	16	O92NX6	O92NX6 rhizobium m
14	7	4.3	173	8	O9XKN3	O9XKN3 anthophora
15	7	4.3	179	10	O9MAZ1	O9MAZ1 glycine max
16	7	4.3	181	17	O27963	O27963 archaeeoglob

17	7	4.3	196	10	O81421
18	7	4.3	197	16	O92L50
19	7	4.3	217	13	O9DDT8
20	7	4.3	227	10	O9LEI4
21	7	4.3	229	10	O9LEI5
22	7	4.3	229	10	O948V4
23	7	4.3	233	10	O9AVU1
24	7	4.3	238	5	O19609
25	7	4.3	238	10	O81258
26	7	4.3	240	2	P94207
27	7	4.3	276	16	O92J54
28	7	4.3	298	16	O9KCS6
29	7	4.3	313	16	O929F3
30	7	4.3	319	10	O9M2J6
31	7	4.3	329	10	O9SUT6
32	7	4.3	341	10	O9FKO5
33	7	4.3	352	10	O9SLJ9
34	7	4.3	363	10	O9L148
35	7	4.3	382	3	O9P5J2
36	7	4.3	391	16	O9A2D4
37	7	4.3	397	5	O94532
38	7	4.3	404	13	O9DFB0
39	7	4.3	424	11	O9JMB0
40	7	4.3	425	5	O9S86
41	7	4.3	426	5	O9VLK0
42	7	4.3	440	5	O9XX21
43	7	4.3	464	5	O9W265
44	7	4.3	480	10	O94CK0
45	7	4.3	491	10	O9FKX5

ALIGNMENTS

RESULT 1
ID O9GZM0 PRELIMINARY; PRT; 162 AA.
AC O9GZM0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE DJ604K5.1 (15 KDA SELENOPROTEIN).
GN DJ604K5.1 OR SEP15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearce A.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed=10945981;
RA Kumaraswamy E., Malynk A., Korotkov K.V., Kozayavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.*;
RT *Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.
DR U. Biol. Chem. 275:35540-35547(2000).
DR EMBL; AL121989; CAC04186.1; -.
DR EMBL; AF288991; AAG31556.1; -.
SQ SEQUENCE 162 AA; 17750 MW; 5B94FBB423A089EE CRC64;

Query Match 90.1%; Score 146; DB 4; Length 162;
Best local Similarity 100.0%; Pred. No. 6.6e-142;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 RLILATVILQAVSAFGAFSSSEACRELGFSNNLCSSCDLIGQNLQIDPDRCGCCQEEA 75
DB 16 RLILATVILQAVSAFGAFSSSEACRELGFSNNLCSSCDLIGQNLQIDPDRCGCCQEEA 75

QY 76 QEFETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 135
DB 76 QEFETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 135
QY 136 NIAEELSLKWNDSVEEFLSEKLERI 162
DB 136 NIAEELSLKWNDSVEEFLSEKLERI 162

RESULT 2

Q9ERR7 PRELIMINARY; PRT; 162 AA.
AC Q9ERR7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 15 KDA SELENOPROTEIN.
GN SEP15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519655; Pubmed=10945981;
RA Kumaraswamy E., Malikh A., Korotkov K.V., Kozyavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.
RT POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.";
RL J. Biol. Chem. 275:35540-35547(2000).
DR EMBL: AF288740; AAC31765.1; -;
DR MGD: MGI:1927947; Sep15.
SQ SEQUENCE 162 AA; 17738 MW; 6AB4B37D8C8B7847 CRC64;

Query Match

Best Local Similarity 100.0%; Score 95; DB 11; Length 162;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 CRGCCOEAQETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 126
DB 67 CRGCCOEAQETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 126
QY 127 VLKLLDNGNIAEELSLKWNDSVEEFLSEKLERI 162
DB 127 VLKLLDNGNIAEELSLKWNDSVEEFLSEKLERI 162

RESULT 3

Q923V8 PRELIMINARY; PRT; 162 AA.

AC Q923V8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 15 KDA CYTOSOLIC SELENOPROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Roethlisler D., Kyriakopoulos A., Behne D.;
RT "A 15 kDa selenoprotein in several tissues of the rat."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF390544; AAK73100.1; -;
SQ SEQUENCE 162 AA; 17706 MW; 65E939968C8B72AC CRC64;

Query Match

Best Local Similarity 100.0%; Score 78; DB 11; Length 162;

Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 CRGCCOEAQETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 126
DB 67 CRGCCOEAQETKLIYAGAILLEVCGKXKLGFRPVOVAFVRSDBKPLFRGLQIKYVGSDDLKLLDNG 126
QY 127 VLKLLDNGNIAEELSLK 145
DB 127 VLKLLDNGNIAEELSLK 145

RESULT 4

Q9BS64 PRELIMINARY; PRT; 92 AA.

AC Q9BS64;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 9.7 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW, CHRONIC MYELOGENOUS LEUKEMIA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC005294; AAH05294.1; -;
DR EMBL: BC016359; AAH16359.1; -;
KW Hypothetical protein.
SQ SEQUENCE 92 AA; 9669 MW; 59EABAD57FPAAD58 CRC64;

Query Match

Best Local Similarity 100.0%; Score 77; DB 4; Length 92;

Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 RLILATYIOAVSARGAESSEACRELFSSNLLCSCDLGQFNLLQIDPDCRCCOEEA 75
DB 16 RLILATYIOAVSARGAESSEACRELFSSNLLCSCDLGQFNLLQIDPDCRCCOEEA 75
QY 76 QEFETKLIYAGAILLEVCG 92
DB 76 QEFETKLIYAGAILLEVCG 92

RESULT 5

Q9NAC6 PRELIMINARY; PRT; 152 AA.

AC Q9NAC6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 17.6 KDA PROTEIN.
GN Y76B12C.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; Pubmed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]


```

RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RA  Cordes M.;
RT  "The sequence of C. elegans cosmid Y76B12C.";
RL  Submitted (MAR-2000) to the EMBL/genbank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2;
RA  Waterston R.;
RT  "Direct Submission.";
RL  Submitted (SEP-2001) to the EMBL/genbank/DBJ databases.
DR  EMBL: AC024211; AAF36064.2;
KW  Hypothetical protein.
SQ  SEQUENCE 152 AA; 17593 MW; 9CAA2F9C0ABE5B87 CRC64;

Query Match
Best Local Similarity 5.6%; Score 9; DB 5; Length 152;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  97 ADFVOAFV 105
    |||||
DB  88 RFPVOAFV 96

RESULT 6
O9A1M1  PRELIMINARY; PRT; 97 AA.
AC  O9A1M1:
DT  01-JUN-2001 (TREMBlrel. 17, Created)
DT  01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE  01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE  HYPOTHEICAL PROTEIN SPY0195.
GN  SPY0195.
OS  Streptococcus pyogenes.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_TaxID=1314;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX  MEDLINE=21192684; PubMed=11296296;
RA  Perrelet J.J., Meschan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA  Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA  Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA  Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT  "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RT  Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RL  EMBL: AE006486; AK33292.1;
DR  InterPro: IPR002514; Transposase_8.
DR  Pfam: PF01527; Transposase_8; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 97 AA; 11418 MW; 81F2B17A7D2B60BB CRC64;

Query Match
Best Local Similarity 4.9%; Score 8; DB 16; Length 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  75 ADFETKKL 82
    |||||
DB  64 ADFETKKL 71

RESULT 7
O9VUJ7  PRELIMINARY; PRT; 142 AA.
AC  O9VUJ7:
DT  01-MAY-2000 (TREMBlrel. 13, Created)
DT  01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE  01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE  01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DR  CG7484 PROTEIN.
GN  CG7484.

```

```

OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chapple M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA  Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA  Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RT  Science 287:2185-2195(2000).
RL  EMBL: AE003523; AAF49314.1;
DR  Flybase: FBgn0036745; CG7484.
DR  Pylase: FBgn0036745; CG7484.
SQ  SEQUENCE 142 AA; 15965 MW; 01D06725BB3DB4DF CRC64;

Query Match
Best Local Similarity 4.9%; Score 8; DB 5; Length 142;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  116 LQIKYVRG 123
    |||||
DB  69 LQIKYVRG 76

RESULT 8
O9S578  PRELIMINARY; PRT; 295 AA.
AC  O9S578:
DT  01-MAY-2000 (TREMBlrel. 13, Created)
DT  01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE  01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE  JR2 PROTEIN (PROBABLE BACTERIOPHAGE PROTEIN).
GN  JR2 OR PA0618.
OS  Pseudomonas aeruginosa.
OC  Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC  Pseudomonas.

```

OX NCBI_TaxID=287;
 RN
 RC SEQUENCE FROM N.A.
 RA Nakayama K., Takashima K., Ishihara H., Shinomiya T., Kageyama M.,
 RA Kanaya S., Ohnishi M., Murata T., Terawaki Y., Mori H., Hayashi T.,
 RT "Genetic relationship between bacteriophages and bacteriophages.",
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PAOI;
 RA MEDLINE-20437337; PubMed-10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Telenko E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lapidis K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 RT opportunistic pathogen";
 RL Nature 406:959-964(2000).
 DR EMBL; AB030825; BAA83156.1; -;
 DR EMBL; AE004498; AAC04007.1; -;
 KM Complete proteome.
 SQ SEQUENCE 295 AA; 31958 MW; 44444BC3D92CDE CRC64;

Query Match 4.9%; Score 8; DB 16; Length 295;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 SDPVKLL 131
 |||||
 DB 44 SDPVKLL 51

RESULT 9
 O9UKU7 PRELIMINARY; PRT; 365 AA.
 AC O9UKU7;
 DT 01-OCR-2000 (TREMBlrel. 15, Created)
 DT 01-OCR-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCR-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOTHETICAL 40.2 KDA PROTEIN (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-CS7BL/6J;
 RA MEDLINE-2015471; PubMed-10693804;
 RA Millonig J.H., Millen K.J., Hatten M.E.,
 RT "The mouse Dreher gene *lmx1a* controls formation of the roof plate in
 RT the vertebrate CNS";
 RL Nature 403:764-769(2000).
 DR EMBL; AF226663; AAF43013.1; -;
 KM Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 365 AA; 40209 MW; 86329D3A8BA5C59F CRC64;

Query Match 4.9%; Score 8; DB 11; Length 365;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ATVLQAVS 27
 |||||
 DB 180 ATVLQAVS 187

RESULT 10
 O9WK13

ID O9WK13 PRELIMINARY; PRT; 641 AA.
 AC O9WK13;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE ISOMALTOTRIO-DEXTRANASE PRECURSOR PRECURSOR (EC 3.2.1.95).
 GN DEXT.
 OS Brevibacterium fuscum var. dextranolyticum.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Brevibacterium.
 OC Actinomycetales; Micrococciaceae; Brevibacteriaceae; Brevibacterium.
 OX NCBI_TaxID=90748;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-0407;
 RA Mizuno T., Mori H., Ito H., Matsui H., Kimura A., Chiba S.,
 RT "Molecular cloning of isomaltotrio-dextranase gene from *Brevibacterium*
 RT fuscum

Query Match 4.9%; Score 8; DB 2; Length 641;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAFFSSA 37
 |||||
 DB 119 GAFFSSA 126

RESULT 11
 O96YR9 PRELIMINARY; PRT; 141 AA.
 AC O96YR9;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN S72104.
 GN S72104.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RA PubMed-11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagita M., Nishimura M., Yanagishi A.,
 RA Oshima T., Kikuchi H.,
 RT "Complete genome sequence of an aerobic thermophilic
 RT Crenarchaeon, *Sulfolobus tokodaii* strain7.",
 RL DNA Res. 8:123-140(2001).
 DR EMBL; AP000988; BAB67208.1; -;
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 141 AA; 16616 MW; 000C93003D3EE04D CRC64;

Query Match 4.3%; Score 7; DB 17; Length 141;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 FLSEKLE 160
 |||||
 DB 26 FLSEKLE 32

RESULT 12
 Q9SYL3 PRELIMINARY; PRT; 158 AA.
 AC Q9SYL3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F3F20.17 PROTEIN.
 GN F3F20.17
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 Alfafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 Lucos S., Schwartz J., Shin P., Toriumi M., Vysotskaya V.S.,
 Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AC007153; AAD30621.1; -
 DR InterPro: IPR000345; CYTC_heme_bind.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN.1.
 SQ SEQUENCE 158 AA; 17901 MW; A181CB37DA433B8B CRC64;

Query Match 4.3%; Score 7; DB 10; Length 158;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GALEVC 91
 |||||
 DB 79 GALEVC 85

RESULT 13
 Q92NX6 PRELIMINARY; PRT; 159 AA.
 AC Q92NX6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE PUTATIVE RNA POLYMERASE SIGMA FACTOR PFCI PROTEIN.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21368334; PubMed=11474104;
 RA Gallibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
 Barloy-Hubier F., Barnett M.J., Becker A., Boisdard P., Bothe G.,
 Boutary M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
 Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 Gloux S., Godle T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
 Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
 Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
 Masny D., Palm C., Peck M.C., Pohl T.M., Portetlelle D., Purnelle B.,
 Rampeger U., Surzycki R., Thebaud P., Vandebol M.,
 Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
 RA "The composite genome of the legume symbiont Sinorhizobium meliloti.";
 RT Science 293:668-672(2001).
 DR EMBL: AL591789; CAC4610.1; -
 KW Complete proteome.
 SQ SEQUENCE 159 AA; 18395 MW; 5E3A1654D55C3010 CRC64;

Query Match 4.3%; Score 7; DB 16; Length 159;

Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 137 IAEELSI 143
 |||||
 DB 129 IAEELSI 135

RESULT 14
 Q9XKN3 PRELIMINARY; PRT; 173 AA.
 AC Q9XKN3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 GN CYTB.
 OS Anthophora acervorum.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Apoidea; Apidae; Anthophora.
 OX NCBI_TaxID=60897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Roulianos S., Schmid-Hempel R., Roubik D., Schmid-Hempel P.;
 RT "Relationships within the Apinae.";
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
 COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
 COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
 CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
 BOUND TO THE PROTEIN (BY SIMILARITY).
 CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
 CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
 DR EMBL: AF002731; AAD29081.1; -
 DR InterPro: IPR000179; Cyt_b_b6.
 DR Pfam: PF00033; cytochrome_b_N; 1.
 KW Electron transport; Heme; Mitochondrion; Respiratory chain;
 KM Transmembrane.
 FT NON_TER 1
 FT NON_TER 173
 SQ SEQUENCE 173 AA; 20314 MW; C2BA74FEA9F40FC1 CRC64;

Query Match 4.3%; Score 7; DB 8; Length 173;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
 |||||
 DB 121 LGFSSNL 127

RESULT 15
 Q9MAZ1 PRELIMINARY; PRT; 179 AA.
 AC Q9MAZ1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE NONCATALYTIC COAT PROTEIN ZETA2-COP.
 GN COPZ2.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hahn Y., Chung J.H.;

"Identification of zeta-COP genes from various organisms."
RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB040543; BAA93004.1; -
DR InterPro; IPR000804; Clat_adaptor_s.
DR Pfam; PF01217; Clat_adaptor_s; 1.
KM Coat protein.
SQ SEQUENCE 179 AA; 19790 MW; CC01CFE2EB84341C CRC64;

Query Match 4.3%; Score 7; DB 10; Length 179;
Best Local Similarity 100.0%; Pred.No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 LATVLOA 25
|||
Db 90 LATVLOA 96

Search completed: September 18, 2002, 16:38:58
Job time: 213 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:33:55 ; Search time 18.81 Seconds
(without alignments)
827.564 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 162

Sequence: 1 MAAGPSCGLVPAFGKRLLLA.....ILKMTDSVEEFLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR.71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.9	295	2	T44541
2	8	4.9	641	2	JC7142
3	8	4.9	743	2	E83728
4	8	4.9	878	2	AB0478
5	7	4.3	83	1	S25307
6	7	4.3	145	2	S10027
7	7	4.3	158	2	E86191
8	7	4.3	173	2	S10975
9	7	4.3	181	2	A69540
10	7	4.3	196	2	T01425
11	7	4.3	197	2	T26829
12	7	4.3	197	2	C88809
13	7	4.3	282	2	AE2618
14	7	4.3	290	2	D97400
15	7	4.3	298	2	E83836
16	7	4.3	313	2	AG1722
17	7	4.3	319	2	T15980
18	7	4.3	320	2	S32966
19	7	4.3	329	2	T13016
20	7	4.3	352	2	C84603
21	7	4.3	363	2	T49209
22	7	4.3	365	1	MNRR4H
23	7	4.3	382	1	T49724
24	7	4.3	390	1	AA2724
25	7	4.3	390	2	AG2463
26	7	4.3	391	2	E87699
27	7	4.3	396	2	H82728
28	7	4.3	412	1	ITMSC
29	7	4.3	424	2	A50000

30	7	4.3	455	2	T26730	hypothetical prote
31	7	4.3	469	2	D65058	hypothetical prote
32	7	4.3	478	2	D64230	NADH oxidase (nox)
33	7	4.3	491	2	B86155	probable polygalac
34	7	4.3	505	2	T35107	hypothetical prote
35	7	4.3	507	2	B69957	conserved hypothet
36	7	4.3	510	2	A53802	N-acetylglactosam
37	7	4.3	512	2	AF3402	exopolysphatase
38	7	4.3	516	2	T33061	hypothetical prote
39	7	4.3	600	2	B84546	probable splicing
40	7	4.3	627	2	D71729	dnar-type molecula
41	7	4.3	627	2	A97729	dnar protein (limp
42	7	4.3	660	2	S71949	metalloproteinase
43	7	4.3	667	2	S74254	homeotic protein s
44	7	4.3	674	2	T40214	hypothetical prote
45	7	4.3	770	2	T01527	hypothetical prote

ALIGNMENTS

RESULT 1
T44541
hypothetical protein JR2 [imported] - Pseudomonas aeruginosa
N:Alternate names: probable bacteriophage protein PA0618
C:Species: Pseudomonas aeruginosa
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 31-Dec-2000
C:Accession: T44541; H83566
R:Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.;
submitted to the EMBL Data Library, August 1999
A:Description: Genetic relationship between bacteriocius and bacteriophages.
A:Reference number: 222790
A:Accession: T44541
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-295 <NAK>
A:Cross-References: EMBL:AB030825; PIDN:BA083156.1
A:Experimental source: strain PA01
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L
; Lory, S.; Olson, M.V.
N:ture 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: H83566
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-References: GB:AE004498; GB:AE004091; NID:g9946491; PIDN:AA04007.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0618
A>Note: JR2

Query Match 4.9%; Score 8; DB 2; Length 295;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 SDPVKLL 131
DB 44 SDPVKLL 51

RESULT 2
JC7142
dextranase (EC 3.2.1.11) - Brevibacterium fuscum var. dextranolyticum
N:Alternate names: isomaltotri-o-dextranase
C:Species: Brevibacterium fuscum var. dextranolyticum
C>Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C:Accession: JC7142; PC7042
R:Mizuno, T.; Mori, H.; Ito, H.; Matsui, H.; Kimura, A.; Chiba, S.
Biosci. Biotechnol. Biochem. 63, 1582-1588, 1999

A:Title: Molecular cloning of isomaltotriose-dextranase gene from *Brevibacterium fuscum* v4
 A:Reference number: JC7142; MUID:20008261
 A:Accession: JC7142
 A:Molecule type: DNA
 A:Residues: 1-641 <MIZ>
 A:Cross-references: DDBJ:AB025195
 A:Experimental source: strain 0407
 A:Accession: PC7042
 A:Molecule type: protein
 A:Residues: 38-56;98-109;141-147;149-169;213-231;298-307;309-321;323-342;446-458;497-515
 C:Genetics:
 A:Gene: dext
 C:Keywords: glycosidase; hydrolase

Query Match 4.9%; Score 8; DB 2; Length 641;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAFFSEA 37
 DB 119 GAFFSEA 126

RESULT 3
 E83728
 phosphoribosylformylglycinamide synthetase I purO [imported] - *Bacillus halodurans* (st
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: E83728
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hird
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: E83728
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-743 <STO>
 A:Cross-references: GB:AP001509; GB:BA000004; MID:g10173176; PIDN:BAB04348.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: purO
 C:Superfamily: phosphoribosylformylglycinamide synthetase component II

Query Match 4.9%; Score 8; DB 2; Length 743;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACRELG 42
 DB 525 SEACRELG 532

RESULT 4
 AE0478
 phosphoenolpyruvate carboxylase (EC 4.1.1.31) [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AE0478
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Rariga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davey, P.; Dougan, G.;
 11. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AE0478
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-878 <RUR>
 A:Cross-references: GB:AL590842; PIDN:CAC93393.1; PID:g15981839; GSPDB:GN00175
 C:Genetics:
 A:Gene: ppc

C:Superfamily: phosphoenolpyruvate carboxylase
 C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 4.9%; Score 8; DB 2; Length 878;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 82 LYAGALE 89
 DB 632 LYAGALE 639

RESULT 5
 S25307
 probable allophycocyanin linker protein - red alga (*Cyanidium caldarium*) chloroplast
 C:Species: chloroplast *Cyanidium caldarium*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S25307
 R:Valentin, K.; Mäld, U.; Emlich, A.; Zetsche, K.
 Plant Mol. Biol. 20, 267-276, 1992
 A:Title: Organization and expression of a phycobilliprotein gene cluster from the unic
 A:Reference number: S25306; MUID:93004479
 A:Accession: S25307
 A:Molecule type: DNA
 A:Residues: 1-83 <VAL>
 A:Cross-references: EMBL:X57251; MID:g17969; PIDN:CAA40532.1; PID:g17971
 C:Genetics:
 A:Gene: apcL
 A:Superfamily: probable allophycocyanin linker protein
 C:Keywords: chloroplast

Query Match 4.3%; Score 7; DB 1; Length 83;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 FSSNMLC 49
 DB 13 FSSNMLC 19

RESULT 6
 S10027
 alpha-amylase inhibitor, tetrameric, chain CMI precursor - wheat
 C:Species: *Triticum aestivum* (common wheat)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 20-Jun-2000
 C:Accession: S10027
 R:Garcia-Maroto, F.; Marana, C.; Mena, M.; Garcia-Olmedo, F.; Carbonero, P.
 Plant Mol. Biol. 14, 845-853, 1990
 A:Title: Cloning of cDNA and chromosomal location of genes encoding the three types o
 A:Reference number: S10027; MUID:91346675
 A:Accession: S10027
 A:Molecule type: mRNA
 A:Residues: 1-145 <GAR>
 A:Cross-references: EMBL:X17575; MID:g21700; PIDN:CAA35598.1; PID:g21701
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: alpha-amylase inhibitor; tetramer
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-145/Product: alpha-amylase inhibitor chain CMI #status predicted <MAT>

Query Match 4.3%; Score 7; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LLATVL 23
 DB 10 LLATVL 16

RESULT 7

E86191
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86191
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso, Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E86191
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-158 <STO>
 A:Cross-references: GB:AE005172; NID:g4836919; PIDN:ADJ0621.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 4.3%; Score 7; DB 2; Length 158;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 GATLEVC 91
 |||||
 DB 79 GATLEVC 85

RESULT 8
 S10975
 T-cell surface glycoprotein CD3 delta chain precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
 C:Accession: S10975
 R: Davies, J.D.; Mueller, D.; Wilson, D.B.; Gold, D.P.
 Nucleic Acids Res. 18, 4617, 1990
 A:Title: Nucleotide sequence of a cDNA encoding the rat T3 delta chain.
 A:Reference number: S10975; MUID:90356424
 A:Accession: S10975
 A:Molecule type: mRNA
 A:Residues: 1-173 <DAV>
 A:Cross-references: EMBL:X53430; NID:g55913; PIDN:CAA37521.1; PID:g55914
 C:Superfamily: T-cell surface glycoprotein CD3 delta chain; Immunoglobulin homology
 C:Keywords: glycoprotein; transmembrane protein
 F:30-75/Domain: Immunoglobulin homology <IMM>

Query Match 4.3%; Score 7; DB 2; Length 173;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LLLATVL 23
 |||||
 DB 10 LLLATVL 16

RESULT 9
 A69540
 conserved hypothetical protein AF2321 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: A69540
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Atliach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
 A:Reference number: A69250; MUID:98049343
 A:Accession: A69540
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-181 <KLE>
 A:Cross-references: GB:AE000944; GB:AE000782; NID:g2689267; PIDN:AAB88935.1; PID:g264

Query Match 4.3%; Score 7; DB 2; Length 181;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACREL 41
 |||||
 DB 77 SEACREL 83

RESULT 10
 T01425
 hypothetical protein T2H3.15 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Mar-1999
 C:Accession: T01425
 R: Kalicki, J.; Gibson, A.
 submitted to the EMBL Data Library, August 1998
 A:Description: The sequence of A. thaliana T2H3.
 A:Reference number: Z14324
 A:Accession: T01425
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-196 <KAL>
 A:Cross-references: EMBL:AF075597; NID:g3298610; PID:g3377809
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4
 A:introns: 77/1
 A:Note: T2H3.15

Query Match 4.3%; Score 7; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 138 AEBLSIL 144
 |||||
 DB 29 AEBLSIL 35

RESULT 11
 T26829
 hypothetical protein Y43C5B.1 - Caenorhabditis elegans (fragment)
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26829
 R:White, S.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: Z20273
 A:Accession: T26829
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-197 <WIL>
 A:Cross-references: EMBL:AL023841; PIDN:CAA19512.1; GSPDB:GN00022; CESP:Y43C5B.1
 A:Experimental source: clone Y43C5B
 C:Genetics:
 A:Gene: CESP:Y43C5B.1
 A:Map position: 4

Query Match 4.3%; Score 7; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 FSSSEACR 39
 |||||
 DB 18 FSSSEACR 24

RESULT 12

C88809
 Protein Y43C5B.1 [Imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: C88809
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_elegans/
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: C88809
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-197 <STO>
 A:Cross-references: GB:chr_IV; PIDN:CA19512.1; PID:g380944; GSPDB:GN00022; CESP:Y43C5B
 C:Genetics:
 A:Gene: Y43C5B.1
 A:Map position: 4

Query Match 4.3%; Score 7; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 FSSSEACR 39
 |||||
 DB 18 FSSSEACR 24

RESULT 13

AE2618
 rRNA methylase [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
 C:Accession: AE2618
 R:Mod, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McClellan,
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AE2618
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-282 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AA141363.1; PID:g17738678; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: trmH
 A:Map position: circular chromosome
 C:Superfamily: conserved hypothetical protein HI0860

Query Match 4.3%; Score 7; DB 2; Length 282;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SVEEFLS 156
 |||||
 DB 188 SVEEFLS 194

RESULT 14

D97400
 rRNA methylase (AP001517) [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: D97400
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: D97400
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-290 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86157.1; PID:g15155248; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_593
 A:Map position: circular chromosome
 C:Superfamily: conserved hypothetical protein HI0860

Query Match 4.3%; Score 7; DB 2; Length 290;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 SVEEFLS 156
 |||||
 DB 196 SVEEFLS 202

RESULT 15

E83836
 reductase BH1493 [Imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: E83836
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: E83836
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-298 <STO>
 A:Cross-references: GB:AP001512; GB:BA000004; MID:g10174030; PIDN:BA05212.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH1493

Query Match 4.3%; Score 7; DB 2; Length 298;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 KRLLIAT 21
 |||||
 DB 100 KRLLIAT 106

Search completed: September 18, 2002, 16:38:13
 Job time: 258 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:36:11 ; Search time 10.7 Seconds

(without alignments)
586.221 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 162

Sequence: 1 MAAGPSGCLVPAFGKRLLLA.....ILKWNDSVEEFLSEKLERI 162

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	47.5	162	1 SE15_HUMAN	O60613 homo sapien
2	69	42.6	162	1 SE15_MOUSE	O96877 mus musculu
3	69	42.6	162	1 SE15_RAT	O923V8 rattus norv
4	8	4.9	640	1 DEXT_ARTGO	P70744 arthrobacte
5	8	4.9	743	1 PURIL_BACHD	O9Kf57 bacillus ha
6	8	4.3	83	1 YC20_GALSD	P48409 tridictoria s
7	7	4.3	145	1 IAO1_WHEAT	P16850 triticum ae
8	7	4.3	165	1 URI_ONCMY	O93448 oncorhynchu
9	7	4.3	173	1 C3D_RAT	P19377 rattus norv
10	7	4.3	210	1 KAB2_OLDAF	P58454 oldelandia
11	7	4.3	320	1 YB9K_YEAST	P38342 saccharomyc
12	7	4.3	365	1 VNS2_AHSY9	P27279 african hor
13	7	4.3	390	1 RPSD_ANASP	P26683 anabaena sp
14	7	4.3	396	1 DXR_XYLPA	O9pe10 xyliella fas
15	7	4.3	412	1 ALAT_MUSCR	P26595 mus caroli
16	7	4.3	433	1 HEM2_SPIOL	P24493 spinacia ol
17	7	4.3	445	1 YGCS_ECOLI	O46909 escherichia
18	7	4.3	478	1 NAOX_MYGE	O46908 mycoplasma
19	7	4.3	507	1 YOGP_BAGSU	P54493 bacillus su
20	7	4.3	510	1 BAGT_MOUSE	O09199 mus musculu
21	7	4.3	627	1 DNAK_RICCN	O92136 rickettsia
22	7	4.3	667	1 DNAX_MOUSE	O92dx9 rickettsia
23	7	4.3	674	1 SIX5_MOUSE	P70178 mus musculu
24	7	4.3	674	1 CWF4_SCHPO	P87312 mus musculu
25	7	4.3	830	1 GYBA_CLOAB	P94605 clostridium
26	7	4.3	831	1 YPDD_ECOLI	P77439 escherichia
27	7	4.3	1284	1 NRK4_DROME	O94887 drosophila
28	7	4.3	1450	1 CFTR_RABIT	O00554 oryctolagus
29	7	4.3	1476	1 CFTR_MOUSE	P26361 mus musculu
30	7	4.3	1480	1 CFTR_HUMAN	P13569 homo sapien
31	7	4.3	1481	1 CFTR_BOVIN	P35071 bos taurus
32	7	4.3	1481	1 CFTR_SHEEP	O00555 ovnis arles
33	6	3.7	79	1 Y3K_HCMVA	P21600 human cytom

34	6	3.7	85	1 YQAH_BACSU	P45905 bacillus su
35	6	3.7	90	1 REV_HV1ZH	P05868 human immun
36	6	3.7	95	1 YCNE_BACSU	P94425 bacillus su
37	6	3.7	100	1 R14_CYPAP	P48138 cyanophora
38	6	3.7	119	1 R18E_SULSO	P95990 sulfolobus
39	6	3.7	121	1 SRI4_ARATH	O04421 arabidopsis
40	6	3.7	129	1 LYC_EQUAS	P11375 equus asinu
41	6	3.7	131	1 AGSM_MOUSE	O03288 mus musculu
42	6	3.7	131	1 AGSM_VULVU	P79407 vulpes vulp
43	6	3.7	131	1 ATPE_GUTH	O78492 guillardia
44	6	3.7	131	1 SVR_BUCAP	O08888 buchnera ap
45	6	3.7	132	1 AGSM_HUMAN	P42127 homo sapien

ALIGNMENTS

```

RESULT 1
ID SE15_HUMAN STANDARD; PRT; 162 AA.
AC O60613; O9NR01; O9BS64; O9GZM0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEP15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 98-106; 123-127 AND 146-158.
RX MEDLINE-98204881; PubMed-9535873;
RA Gladyshev V.N., Jeang K.-T., Wootton J.C., Hatfield D.L.;
RT "A new human selenium-containing protein. Purification,
RT characterization, and cDNA sequence."
RL J. Biol. Chem. 273:8910-8915 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20519655; PubMed-10945981;
RA Kumaraswamy E., Malykh A., Korotkov R.V., Kozayavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology."
RL J. Biol. Chem. 275:35540-35547 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Ryu M., Moon E.;
RT "The human 15-kDa selenoprotein gene: characterisation of the genomic
RT structure and functional analysis of the promoter."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Pearce A.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Tissue-Bone marrow, and kidney;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SUBUNIT: SEEMS TO BE EITHER OLIGOMERIC OR HETEROOLIGOMERIC.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- MASS SPECTROMETRY: MW=14870; METHOD=Electrospray; RANGE=7-152.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC

```

```

DR EMBL: AF051894; AAC15478.1; -
DR EMBL: AF288891; AAC31556.1; -
DR EMBL: AF288992; AAC31557.1; -
DR EMBL: AF267986; AAF78966.1; ALT_SEQ.
DR EMBL: AF267982; AAF78966.1; JOINED.
DR EMBL: AF267983; AAF78966.1; JOINED.
DR EMBL: AF267984; AAF78966.1; JOINED.
DR EMBL: AF267985; AAF78966.1; JOINED.
DR EMBL: AL121989; CAC04186.1; ALT_SEQ.
DR EMBL: BC005294; AAH05294.1; ALT_TERM.
DR EMBL: BC016359; AAH16359.1; ALT_TERM.
DR MIM: 606254; -
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPTOIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17743 MW; 463EACBA23B4CDBE CRC64;

Query Match 47.5%; Score 77; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 6.3e-71;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 RLATVIAVQAASARGAFSSACRELGSSNLCSGDLGQFNLDLPDRCGCCOEAA 75
DB 16 RLATVIAVQAASARGAFSSACRELGSSNLCSGDLGQFNLDLPDRCGCCOEAA 75
QY 76 QFETKRLVAGAILVEVC 92
DB 76 QFETKRLVAGAILVEVC 92

RESULT 2
SEIS_MOUSE STANDARD; PRT; 162 AA.
AC QSER7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=20519655; PubMed=10945981;
RA Kumaraswamy E., Malikh A., Korotkov A.V., Kozayavkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kDa selenoprotein gene.
RT Possible role of the protein in cancer etiology.";
RL J. Biol. Chem. 275:35540-35547(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF288740; AAC31765.1; ALT_SEQ.
DR MGD; MGI:1927947; Sep15.
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPTOIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17731 MW; 771E047D8C9F3C17 CRC64;

Query Match 42.6%; Score 69; DB 1; Length 162;

```

```

Best Local Similarity 100.0%; Pred. No. 8.6e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 KLGRFPQVAQVRSRDKPLFRGLQIKYVGSDPVLKLLDNGNIAEELSTIKWNTDSYEE 153
DB 94 KLGRFPQVAQVRSRDKPLFRGLQIKYVGSDPVLKLLDNGNIAEELSTIKWNTDSYEE 153
QY 154 FLSEKLERI 162
DB 154 FLSEKLERI 162

RESULT 3
SEIS_RAT STANDARD; PRT; 162 AA.
AC Q923V8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEPI5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RX SEQUENCE FROM N.A.
RA Roethlein D., Kyriakopoulos A., Behne D.;
RT "A 15 kDa-selenoprotein in several tissues of the rat.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF390544; AAR73100.1; ALT_SEQ.
KW Selenium; Selenocysteine; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPTOIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17775 MW; 73985E968C9F36FC CRC64;

Query Match 42.6%; Score 69; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 8.6e-63;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 KLGRFPQVAQVRSRDKPLFRGLQIKYVGSDPVLKLLDNGNIAEELSTIKWNTDSYEE 153
DB 94 KLGRFPQVAQVRSRDKPLFRGLQIKYVGSDPVLKLLDNGNIAEELSTIKWNTDSYEE 153
QY 154 FLSEKLERI 162
DB 154 FLSEKLERI 162

RESULT 4
SEIS_ARTGO STANDARD; PRT; 640 AA.
AC P70744;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-
DE glucanohydrolase) (Endodextrinase).
OS Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1665;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T-3044;
RA Oguma T., Kurokawa T., Tobe K., Kitao S., Kobayashi M.;
RL Submitted (Nov-1996) to the EMBL/Genbank/DDJ databases.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic
CC linkages in dextran.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 49 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D88361; BA013596.1; -
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 32
FT CHAIN 33 640
FT DEXTRANASE.
SQ SEQUENCE 640 AA; 71142 MW; 4AC7864A96A537D2 CRC64;

Query Match 4.9%; Score 8; DB 1; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GAEFSSEA 37
DB 125 GAEFSSEA 132

RESULT 5
PRRL_BACCHD STANDARD; PRT; 743 AA.
AC 09KRF57;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
GN PURL OR BH0629.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=8665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
CC glutamine + H(2)O = ADP + phosphate + 5'-
CC phosphoribosylformylglycinamide + L-glutamate.
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNTS; PURO AND PURL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; AP001509; BAB04348.1; -
DR InterPro; IPR000728; AIRS_related.
DR Pfam; PF00586; AIRS; 2.
DR Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
FT NP_BIND 111 122
FT ATP (POTENTIAL).
SQ SEQUENCE 743 AA; 80008 MW; 005FC0855D20D84D CRC64;

Query Match 4.9%; Score 8; DB 1; Length 743;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 SEACRELG 42
DB 525 SEACRELG 532

RESULT 6
YC20_GALSU STANDARD; PRT; 83 AA.
ID YC20_GALSU
AC P48A09;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 9.5 kDa protein ycf20.
GN YCF20.
OS Caldريا sulphuraria.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Eukarya.
OX NCBI_TaxID=130081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-14-1-1 / ISOLATE 107.79/GOETTGENG;
RX MEDLINE=93004479; PubMed=1391770;
RA Valentin K., Maid U., Emich A., Zetsche K.;
RT "Organization and expression of a phycoobiliprotein gene cluster from
RT the unicellular red alga Cyanidium caldarium.";
RL Plant Mol. Biol. 20:267-276(1992).
CC -1- SIMILARITY: BELONGS TO THE YCF20 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57251; CAA0532.1; -
DR PIR; S25307; S25307.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 83 AA; 9485 MW; 990117EC63FE3EF9 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 43 FSSNLIC 49
DB 13 FSSNLIC 19

RESULT 7
IA01_WHEAT STANDARD; PRT; 145 AA.
ID IA01_WHEAT
AC P16850;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)

```

```

DE Alpha-amylase/trypsin inhibitor CMI precursor (Chloroform/methanol-
DE soluble protein CMI).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Triticum.
CX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, CHINESE SPRING; TISSUE=Endosperm;
RX MEDLINE=91346675; Pubmed=2102861;
RA Garcia-Maroto F., Marana C., Mena M., Garcia-Olmedo F., Carbonero P.;
RT "Cloning of cDNA and chromosomal location of genes encoding the three
RT types of subunits of the wheat tetrameric inhibitor of insect alpha-
RT amylase."
RL Plant Mol. Biol. 14:845-853(1990).
RN [2]
RP SEQUENCE OF 26-53.
RC STRAIN=CV, CANDEAL;
RA Barber D., Sanchez-Monge R., Garcia-Olmedo F., Salcedo G., Mendez E.;
RT "Evolutionary implications of sequential homologues among members of
RT the trypsin / alpha-amylase inhibitor family (CM-proteins) in wheat
RT and barley."
RL Biochim. Biophys. Acta 873:147-151(1986).
CC -1- FUNCTION: ALPHA-AMYLASE/TRYPsin INHIBITOR. IT COULD BE INVOLVED
CC IN INSECT DEFENSE MECHANISMS.
CC -1- SUBUNIT: SUBUNIT OF THE TETRAMERIC INHIBITOR.
CC -1- TISSUE SPECIFICITY: ENDOSPERM.
CC -1- MISCELLANEOUS: CM PROTEINS WOULD BE INVOLVED IN THE COOKING
CC QUALITY OF PASTA.
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPsin/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17575; CAA35598.1; -.
DR PIR: A25310; A25310.
DR PIR: S10027; S10027.
DR HSP: P01087; 181U.
DR InterPro: IPR003612; AAI.
DR InterPro: IPR001768; Cereal_tryp_aml1_inh.
DR Pfam: PF00234; tryp_alpha_aml1; 1.
DR PRINTS: PRO0808; AMYLASEINHTR.
DR SMART: SMO0499; AAI; 1.
DR PROSITE: PS00426; CEREAL_TRYP_AML1_INH; 1.
KW Serine protease inhibitor; Alpha-amylase inhibitor; Multigene family;
KW Seed; Signal.
FT SIGNAL 1 25
FT CHAIN 26 145 ALPHA-AMYLASE/TRYPsin INHIBITOR CMI.
FT FT 145
SQ SEQUENCE 145 AA; 15517 MW; 119EDBD8C1229672 CRC64;

```

```

Query Match 4.3%; Score 7; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 17 LLLATVL 23
DB 10 LLLATVL 16

```

```

RESULT 8
ID URL_ONCMY STANDARD; PRT; 165 AA.
AC 093448;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Urotensin I precursor.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
CX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hypothalamus;
RA Barsyte D., Tipping D., Brennan J., Baker B., Lovejoy D.;
RT "Cloning of rainbow trout urotensin-I."
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
CC SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CC CORTICOTROPIN-RELEASING FACTOR. THE NONHORMONAL PORTION OF THIS
CC PRECURSOR MAY BE A UROTENSIN BINDING PROTEIN, UROPHYSIN.
CC -1- SIMILARITY: BELONGS TO THE SAVIAGINE/CORTICOTROPIN-RELEASING
CC FACTOR/UROTENSIN I FAMILY OF PEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ005264; CAA06461.1; -.
DR InterPro: IPR00187; CRF.
DR InterPro: IPR003620; Urocortin_CRF.
DR PIR: PF00473; CRF; 1.
DR PRODOM: PD005970; Urocortin_CRF; 1.
DR SMART: SM00039; CRF; 1.
DR PROSITE: PS00511; CRF; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Signal.
FT SIGNAL 1 18
FT CHAIN 19 120 UROPHYSIN (POTENTIAL).
FT PEPTIDE 123 163 UROPHYSIN-I.
FT MOD_RSS 163 163 AMIDATION (G-164 PROVIDE AMIDE GROUP) (BY
FT FT 163
SQ SEQUENCE 165 AA; 18631 MW; 3263357EE7653A1A CRC64;

```

```

Query Match 4.3%; Score 7; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 17 LLLATVL 23
DB 8 LLLATVL 14

```

```

RESULT 9
ID CD3D_RAT STANDARD; PRT; 173 AA.
AC P19377;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD3 delta chain precursor (T-cell receptor
DE T3 delta chain).
GN CD3D OR T3D.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=90356424; Pubmed=2143819;
RA Davies J.D., Mueller D., Wilson D.B., Gold D.P.;
RT "Nucleotide sequence of a cDNA encoding the rat T3 delta chain.";

```


CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch)

 DR EMBL; X70529; CAA49930.1; -
 DR EMBL; Z36134; CAA85228.1; -
 DR PIR; S32966; S32966.
 DR SGD; S0000469; TSC10.
 DR InterPro; IPR002198; ADH_short.
 DR Pfam; PF00106; adh_short; 1.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 FT TRANSMEM 280 300 POTENTIAL.
 SQ SEQUENCE 320 AA; 35986 MW; C5F3E004082FD242 CRC64;

 Query Match 4.3%; Score 7; DB 1; Length 320;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 110 PKLFRGL 116
 |||||
 Db 122 PKLFRGL 128

 RESULT 12
 VNS2_AHSV9 STANDARD; PRT; 365 AA.
 ID VNS2_AHSV9
 AC P27279;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Nonstructural protein NS2.
 GN S8.
 OS African horse sickness virus 9 (AHSV-9) (African horse sickness virus
 (serotype 9)).
 OC Viruses; dsRNA viruses; Reoviridae; Orbiviruses.
 OX NCBI_TaxID=10897;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92024120; PubMed=1656603;
 RA Van Staden V., Theron J., Greyling B.J., Huismans H., Nel L.H.;
 RT "A comparison of the nucleotide sequences of cognate NS2 genes of
 RT three different orbiviruses.";
 RL Virology 185:500-504(1991).
 CC -1- FUNCTION: SSRNA-BINDING PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M69090; AAA42540.1; -
 DR PIR; A40788; MNXRAH.
 KM Nonstructural protein; RNA-binding.
 SQ SEQUENCE 365 AA; 41193 MW; 45F3A78754887741 CRC64;

 Query Match 4.3%; Score 7; DB 1; Length 365;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 118 IKYVGRS 124
 |||||
 Db 121 IKYVGRS 127

 RESULT 13
 RPSD_ANASP STANDARD; PRT; 390 AA.
 ID RPSD_ANASP

AC P26683;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE RNA polymerase sigma factor rpoD (Sigma-A).
 GN RPOD OR SIGA OR ALU5263.
 OS Anabaena sp. (strain PCC 7120).
 CC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91193199; PubMed=1901566;
 RA Brahamsa B., Haselkorn R.;
 RT "Isolation and characterization of the gene encoding the principal
 RT sigma factor of the vegetative cell RNA polymerase from the
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL J. Bacteriol. 173:2442-2450(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED. THIS IS THE PRIMARY SIGMA-FACTOR OF THIS
 CC BACTERIA.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M60046; AAA22043.1; -
 DR EMBL; AP003599; BAB76962.1; -
 DR PIR; A42724; A42724.
 DR HSSP; P00579; ISIG.
 DR InterPro; IPR000943; Sigma_70.
 DR Pfam; PF00140; sigma70; 1.
 DR PRINTS; PR00046; SIGMA70FCF.
 DR PROSITE; PS00715; SIGMA70_1; 1.
 DR PROSITE; PS00716; SIGMA70_2; 1.
 KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
 KM DNA-binding; Complete proteome.
 FT DOMAIN 182 195 POLYMERASE CORE BINDING (POTENTIAL).
 FT DNA BIND 351 370 H-T-H MOTIF (BY SIMILARITY).
 SQ SEQUENCE 390 AA; 45641 MW; 604814660815284D CRC64;

 Query Match 4.3%; Score 7; DB 1; Length 390;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 155 LSEKLER 161
 |||||
 Db 120 LSEKLER 126

 RESULT 14
 DXR_XYLFA STANDARD; PRT; 396 AA.
 ID DXR_XYLFA
 AC Q9PEIO;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-DEOXY-2002 (Rel. 41, last annotation update)
 DE 1-deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
 GN reductoisomerase) (1-deoxyxylose-5-phosphate reductoisomerase).
 OS Xylella fastidiosa.
 CC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 CC NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Gardier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Jungblut M.L., Kemper E.L., Kitaajima J.P.,
 RA Krieger J.E., Kuramae E.E., Lalget F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.H.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhami A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesqueto J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasat H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terezzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RL Nature 406:151-159(2000).
 CC -1- FUNCTION: CATALYZES THE NADP-DEPENDENT REARRANGEMENT AND REDUCTION
 OF 1-DEOXY-D-XYULOSE-5-PHOSPHATE (DXP) TO 2-C-METHYL-D-ERYTHRITOL
 4-PHOSPHATE (MEP) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
 = 1-deoxy-D-xylose 5-phosphate + NADPH.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second
 step.
 CC -1- SIMILARITY: BELONGS TO THE DXP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE003942; AAF3858.1;
 DR InterPro: IPR003821; DXP_reductoisomerase.
 DR Pfam: PF02670; DXP_reductoisom: 1
 KW Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT NP_BIND 11 18 NADPH (POTENTIAL).
 SQ SEQUENCE 396 AA; 42076 MW; 14B56FE95748FD97 CRC64;

Query Match 4.3%; Score 7; DB 1; Length 396;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GKRLLA 20
 |||||
 DB 121 GKRLLA 127

RESULT 15
 ID ALAT_MUSCR STANDARD; PRT; 412 AA.
 AC P26595;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Alpha-1-antitrypsin precursor (Alpha-1-antitrypsin) (Alpha-1-
 DE proteinase inhibitor) (AAT).
 OS Mus caroli (Wild mouse) (Ricefield mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10089;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90136592; PubMed=1689000;
 RA Latimer J.J., Berger F.G., Baumann H.;
 RT "Highly conserved upstream regions of the alpha 1-antitrypsin gene in
 RT two mouse species govern liver-specific expression by different
 RT mechanisms.";
 RL Mol. Cell. Biol. 10:760-769(1990).
 CC -1- FUNCTION: INHIBITOR OF SERINE PROTEASES. ITS PRIMARY TARGET IS
 CC ELASTASE, BUT IT ALSO HAS A MODERATE AFFINITY FOR PLASMIN AND
 CC THROMBIN.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: EXPRESSED NOT ONLY IN LIVER BUT ALSO IN KIDNEY
 CC TUBULE CELLS, WHERE IT IS REGULATED BY ANDROGENS DURING
 CC DEVELOPMENT.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M33567; AAA37128.1;
 DR PIR: A34730; ITMSC.
 DR HSSP: P01009; 9API.
 DR InterPro: IPR000215; Serpin.
 DR Pfam: PF00079; serpin.1.
 DR SMART: SM00093; SERPIN; 1.
 DR PROSITE: PS00284; SERPIN; 1.
 KW Serpin; Serine protease inhibitor; Glycoprotein; Plasma; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 412
 FT ACT_SITE 376 377
 FT CARBOHYD 100 100
 FT CARBOHYD 133 133
 FT CARBOHYD 264 264
 FT CARBOHYD 313 313
 SQ SEQUENCE 412 AA; 45872 MW; C1A167063CAF4BD CRC64;

Query Match 4.3%; Score 7; DB 1; Length 412;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ATVLQAV 26
 |||||
 DB 368 ATVLQAV 374

Search completed: September 18, 2002, 16:39:30
 Job time: 199 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:32:50 ; Search time 16.05 Seconds
(without alignments)
246.539 Million cell updates/sec

Title: US-09-676-718A-1
Perfect score: 162
Sequence: 1 MAAGPSGCLVPAFGKRLLA.....ILKNWDSVEEFLSEKLERI 162

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	4.3	346	2	US-08-602-359A-34
2	7	4.3	424	2	US-08-871-268A-23
3	7	4.3	424	3	US-08-871-267B-31
4	7	4.3	424	4	US-09-618-419-31
5	7	4.3	424	4	US-09-163-674-23
6	7	4.3	503	3	US-08-700-651-6
7	7	4.3	510	1	US-08-255-670A-2
8	7	4.3	836	1	US-08-216-971-2
9	7	4.3	836	2	US-08-812-979-2
10	7	4.3	1476	2	US-09-256-703-2
11	7	4.3	1479	2	US-08-951-912-4
12	7	4.3	1479	4	US-09-174-077-4
13	7	4.3	1480	1	US-07-637-621-2
14	7	4.3	1480	1	US-08-136-742A-2
15	7	4.3	1480	1	US-08-135-809A-2
16	7	4.3	1480	1	US-08-466-886-17
17	7	4.3	1480	2	US-08-951-912-2
18	7	4.3	1480	2	US-08-951-912-6
19	7	4.3	1480	2	US-08-469-461-2
20	7	4.3	1480	2	US-08-469-461-4
21	7	4.3	1480	2	US-08-691-605-2
22	7	4.3	1480	2	US-08-453-552A-14
23	7	4.3	1480	3	US-07-890-609-2
24	7	4.3	1480	3	US-07-890-609-4
25	7	4.3	1480	3	US-09-248-026-2
26	7	4.3	1480	4	US-08-469-617-17
27	7	4.3	1480	4	US-08-681-838A-2

28	7	4.3	1480	4	US-08-681-838A-3	Sequence 3, App1
29	7	4.3	1480	4	US-09-174-077-2	Sequence 2, App1
30	7	4.3	1480	4	US-09-174-077-6	Sequence 6, App1
31	7	4.3	1480	5	PCT-US93-11667-2	Sequence 2, App1
32	7	4.3	1480	6	5240846-5	Patent No. 5240846
33	6	3.7	24	3	US-08-964-302A-13	Sequence 13, App1
34	6	3.7	24	4	US-09-441-416A-13	Sequence 13, App1
35	6	3.7	25	1	US-08-366-690-2	Sequence 2, App1
36	6	3.7	28	4	US-08-905-223-328	Sequence 328, App1
37	6	3.7	50	1	US-08-336-343A-16	Sequence 16, App1
38	6	3.7	74	4	US-09-142-565-4	Sequence 4, App1
39	6	3.7	95	4	US-09-142-565-6	Sequence 6, App1
40	6	3.7	99	3	US-09-173-581-8	Sequence 8, App1
41	6	3.7	99	4	US-09-420-915-8	Sequence 8, App1
42	6	3.7	131	3	US-09-102-977-2	Sequence 2, App1
43	6	3.7	131	4	US-09-034-088A-2	Sequence 2, App1
44	6	3.7	132	4	US-09-034-088A-4	Sequence 4, App1
45	6	3.7	135	3	US-09-100-391-12	Sequence 12, App1

ALIGNMENTS

RESULT 1
US-08-602-359A-34
Sequence 34, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MORPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SMANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSHOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HALE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ. ID NO.: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-602-359A-34

Query Match 4.38; Score 7; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 GKRLLA 20
|||||
DB 57 GKRLLA 63

RESULT 2

US-08-871-268A-23
; Sequence 23, Application US/08871268A
; Patent No. 5866391
; GENERAL INFORMATION:
; APPLICANT: Jones, Audrey
; APPLICANT: Cherry, Joel R.
; TITLE OF INVENTION: Aspergillus Porphobilinogen Synthases
; TITLE OF INVENTION: and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866391disk of No. 5866391th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,268A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4809,200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5866391e
US-08-871-268A-23

Query Match 4.38; Score 7; DB 2; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||
DB 25 LGFSSNL 31

RESULT 3

US-08-871-267B-31
; Sequence 31, Application US/08871267B
; Patent No. 6100057
; GENERAL INFORMATION:
; APPLICANT: Elrod, Susan L.
; APPLICANT: Cherry, Joel R.
; APPLICANT: Jones, Audrey
; TITLE OF INVENTION: A Method for Increasing Hemoprotein
; TITLE OF INVENTION: Production in Filamentous Fungi
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6100057o No. 6100057disk of No. 6100057th America, Inc.
; STREET: 405 Lexington Avenue - 64th Fl.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,267B
; FILING DATE: 9-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4771,200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-878-9652
; TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 424 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6100057e
US-08-871-267B-31

Query Match 4.38; Score 7; DB 3; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||
DB 25 LGFSSNL 31

RESULT 4
US-09-618-419-31
; Sequence 31, Application US/09618419
; Patent No. 6261827
; GENERAL INFORMATION:
; APPLICANT: Elrod, Susan L.
; APPLICANT: Cherry, Joel R.
; APPLICANT: Jones, Audrey

TITLE OF INVENTION: A Method for Increasing Hemoprotein
; Production in Filamentous Fungi
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6261827o No. 6261827disk of No. 6261827th America, Inc.
; STREET: 405 Lexington Avenue - 64th Fl.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,419
; FILING DATE: 18-JUL-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871,267
; FILING DATE: 9-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.

REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 4771.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-878-9652
TELEFAX: 212-878-9655
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6261827e
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-618-419-31

Query Match 4.3%; Score 7; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||||
Db 25 LGFSSNL 31

RESULT 5
US-09-163-674-23
Sequence 23, Application US/09163674
Patent No. 6306630
GENERAL INFORMATION:
APPLICANT: Jones, Aubrey
APPLICANT: Cherry, Joel R.
TITLE OF INVENTION: Aspergillus Porphobilinogen Synthases
TITLE OF INVENTION: and Nucleic Acids Encoding Same
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 63066300 No. 6306630disk of No. 6306630th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/163,674
FILING DATE: 30-Sept-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 4809.210-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 424 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6306630e
US-09-163-674-23

Query Match 4.3%; Score 7; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 LGFSSNL 47
|||||||
Db 25 LGFSSNL 31

RESULT 6
US-08-700-651-6
Sequence 6, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEDCH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
TITLE OF INVENTION: INFECTIONS
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 503
TYPE: PRF
ORGANISM: Cryptosporidium parvum
US-08-700-651-6

Query Match 4.3%; Score 7; DB 3; Length 503;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 FLSKLE 160
|||||||
Db 296 FLSKLE 302

RESULT 7
US-08-255-670A-2
Sequence 2, Application US/08255670A
Patent No. 5691180
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B
APPLICANT: SMTH, PETER L.
TITLE OF INVENTION: N-ACETYL-GALACTOSAMINE-TRANSFERASE cDNA
TITLE OF INVENTION: SEQUENCE AND EXPRESSION PRODUCTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/255,670A
FILING DATE: 09-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Obion, No. 5691180man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2363-092-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-255-670A-2

Query Match 4.3%, Score 7, DB 1, Length 510;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 125 DPVTKL 131
|||||
DB 55 DPVTKL 61

RESULT 8
US-08-216-971-2
Sequence 2, Application US/08216971
Patent No. 5639661
GENERAL INFORMATION:
APPLICANT: Welsh, Michael J.
APPLICANT: Sheppard, David N.
TITLE OF INVENTION: NOVEL GENES AND PROTEINS FOR TREATING
TITLE OF INVENTION: CYSTIC FIBROSIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite #510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,971
FILING DATE: 23-MAR-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIZ-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-971-2

Query Match 4.3%, Score 7, DB 1, Length 836;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 155 ISEKLER 161
|||||
DB 49 ISEKLER 55

RESULT 9
US-08-812-979-2

Sequence 2, Application US/08812979
Patent No. 5958693
GENERAL INFORMATION:
APPLICANT: Welsh, Michael J.
APPLICANT: Sheppard, David N.
TITLE OF INVENTION: NOVEL GENES AND PROTEINS FOR TREATING
TITLE OF INVENTION: CYSTIC FIBROSIS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite #510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,979
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,971
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIZ-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-812-979-2

Query Match 4.3%, Score 7, DB 2, Length 836;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 155 ISEKLER 161
|||||
DB 49 ISEKLER 55

RESULT 10
US-09-256-703-2
Sequence 2, Application US/09256703
Patent No. 6294379
GENERAL INFORMATION:
APPLICANT: Dong, Jian-yun
APPLICANT: Kan, Yuet Wai
TITLE OF INVENTION: The Regents of the University of California
TITLE OF INVENTION: Efficient AAV Vectors
FILE REFERENCE: 023070-084910US
CURRENT APPLICATION NUMBER: US/09/256,703
CURRENT FILING DATE: 1999-02-24
PRIOR APPLICATION NUMBER: US 60/075,980
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 1476
TYPE: PRT
ORGANISM: Homo sapiens
OTHER INFORMATION: truncated cystic fibrosis transmembrane

US-09-256-703-2

Query Match 4.3%; Score 7; DB 4; Length 1476;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 11
US-08-951-912-4
; Sequence 4, Application US/08951912
; Patent No. 5972995
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
; TITLE OF INVENTION: FIBROSIS THERAPY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,912
; FILING DATE: 16-OCT-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200116.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1479 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-951-912-4

Query Match 4.3%; Score 7; DB 2; Length 1479;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 12
US-09-174-077-4
; Sequence 4, Application US/09174077
; Patent No. 6329422
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC FIBROSIS THERAPY
; FILE REFERENCE: 200116.403CI
; CURRENT APPLICATION NUMBER: US/09/174,077

; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: US 08/951,912
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1479
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-174-077-4

Query Match 4.3%; Score 7; DB 4; Length 1479;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 13
US-07-637-621-2
; Sequence 2, Application US/07637621
; Patent No. 5407796
; GENERAL INFORMATION:
; APPLICANT: cutting, gary
; APPLICANT: antonarakis, stylianos e
; APPLICANT: kazazian jr., haig h
; TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/637,621
; FILING DATE: 19910104
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: kagan, sarah a
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.030010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
US-07-637-621-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 155 LSEKLER 161
|||||||
DB 49 LSEKLER 55

RESULT 14
US-08-136-742A-2
; Sequence 2, Application US/08136742A
; Patent No. 5670488
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.A., Smith,
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE & RAYMOND
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,742A
; FILING DATE: 02-DEC-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,478
; FILING DATE: 02-DEC-1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Seide, Rochelle K.
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: A30668 (Genzyme Dkt. IG4-9.11)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 765-2519
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-136-742A-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|111111|
DB 49 LSEKLER 55

RESULT 15
US-08-135-809A-2
; Sequence 2, Application US/08135809A
; Patent No. 5688677
; GENERAL INFORMATION:
; APPLICANT: CHENG, SENG H.
; APPLICANT: DITULLIO, PAUL
; APPLICANT: EBBERT, KARL M.
; APPLICANT: MEADE, HARRY M.
; APPLICANT: SMITH, ALAN E.
; TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS CONTAINING
; TITLE OF INVENTION: INACTIVATED HORMONE RESPONSIVE ELEMENTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA

ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/135,809A
; FILING DATE: 13-OCT-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: IG4-9.12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-135-809A-2

Query Match 4.3%; Score 7; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LSEKLER 161
|111111|
DB 49 LSEKLER 55

Search completed: September 18, 2002, 16:37:34
Job time: 284 sec

OM of: US-09-676-718A-1 to: EST: * out_format : pfs

Date: Sep 18, 2002 5:06 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+g2n.model -DEV=x1h  
-O=/cgn2.1/USPTO_spool/US09676718/runat.17092002.143432.21293/app-query.fasta.1.222  
-DB=EST -OEXT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000  
-DELXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODAL=LOCAL -OUTEXT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09676718 @CN1.1.2515  
-NCPR=6 -ICPU=3 -LONGLOG -DEV=TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPRX -WAIT -THREADS=1
```

Search information block:

Query: US-09-676-718A-1

Query length: 162

Database: EST:*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 1625.180000

```
score_list:
Sequence      Strd Orig  ZScore  EScore Len 1 Documentation
gb_est2:BG703377 + 819.00 1699.98 1.9e-85 587 1 BG703377 602685165F1 NIH_MGC_95
gb_est2:BG399386 + 819.00 1699.00 2.2e-85 642 1 BG399386 602441143F1 NIH_MGC_75
gb_est2:BG779736 + 819.00 1699.63 2.3e-85 664 1 BG779736 60268566F1 NIH_MGC_160
gb_est1:AV757896 + 819.00 1698.15 2.4e-85 694 1 AV757896 AV757896 BM Homo sapi
gb_est1:AV718270 + 819.00 1698.05 2.5e-85 700 1 AV718270 AV718270 FHTB Homo sap
gb_est2:BI518529 + 819.00 1697.46 2.6e-85 726 1 BI518529 603061650F1 NIH_MGC_11
gb_est2:BI545979 + 819.00 1697.40 2.7e-85 739 1 BI545979 603188102F1 NIH_MGC_95
gb_est2:BG505091 + 819.00 1697.40 2.7e-85 743 1 BG505091 602551546F1 NIH_MGC_61
gb_est2:BG613877 + 819.00 1697.38 2.7e-85 744 1 BG613877 602639684F1 NIH_MGC_61
gb_est2:BG777365 + 819.00 1697.35 2.7e-85 746 1 BG777365 602664622F1 NIH_MGC_61
gb_est2:BI597212 + 819.00 1697.29 2.7e-85 750 1 BI597212 603350980F1 NIH_MGC_96
gb_est1:AL533548 + 819.00 1696.77 2.9e-85 787 1 AL533548 602549479F1 NIH_MGC_61
gb_est2:BG502649 + 819.00 1696.63 3.0e-85 797 1 BG502649 602549479F1 NIH_MGC_61
gb_est2:BG533149 + 819.00 1696.56 3.0e-85 802 1 BG533149 602580633F1 NIH_MGC_61
gb_est2:BG529329 + 819.00 1696.48 3.0e-85 808 1 BG529329 602585441F1 NIH_MGC_61
gb_est2:BG402494 + 819.00 1696.24 3.1e-85 826 1 BG402494 602465503F1 NIH_MGC_75
gb_est2:BG499649 + 819.00 1696.16 3.1e-85 832 1 BG499649 602545031F1 NIH_MGC_96
gb_est2:BI597109 + 819.00 1695.85 3.2e-85 853 1 BI597109 603251754F1 NIH_MGC_96
gb_est2:BF978924 + 819.00 1695.76 3.3e-85 863 1 BF978924 602149392F2 NIH_MGC_62
gb_est2:BE781717 + 819.00 1695.42 3.4e-85 890 1 BE781717 601467377F1 NIH_MGC_67
gb_est1:AL559824 + 819.00 1694.15 4.1e-85 999 1 AL559824 AL559824 LTI_FL011_B01
gb_est2:BM476636 + 819.00 1693.85 4.2e-85 1026 1 BM476636 AGENCOURT_6476481 NIH
gb_est2:BG503376 + 816.00 1692.35 5.1e-85 664 1 BG503376 602550792F1 NIH_MGC_61
gb_est2:BG612631 + 816.00 1692.27 5.2e-85 669 1 BG612631 602640055F1 NIH_MGC_61
gb_est2:BG613959 + 816.00 1690.62 6.4e-85 778 1 BG613959 602641880F1 NIH_MGC_61
gb_est2:BG613959 + 815.00 1689.77 7.2e-85 699 1 BG613959 602641880F1 NIH_MGC_61
gb_est2:BG107890 + 814.00 1684.32 1.4e-84 803 1 BG107890 602278158F1 NIH_MGC_86
gb_est2:BG530811 + 813.00 1683.94 1.5e-84 807 1 BG530811 602559842F1 NIH_MGC_61
gb_est1:AA305370 + 810.00 1682.25 1.9e-84 531 1 AA305370 EST176439 Colon carcin
gb_est1:AA314273 + 810.00 1682.00 2.4e-84 634 1 AA314273 EST176439 Colon carcin
gb_est1:AV651776 + 810.00 1680.30 2.4e-84 634 1 AV651776 AV651776 GIC Homo sap
gb_est1:AV707435 + 810.00 1679.85 2.5e-84 661 1 AV707435 AV707435 ADB Homo sap
gb_est2:BI668960 + 810.00 1678.22 2.9e-84 719 1 BI668960 603294940F1 NIH_MGC_96
gb_est1:AV710046 + 808.00 1676.22 4.0e-84 628 1 AV710046 AV710046 Cu Homo sap
gb_est1:BM630212 + 806.00 1674.76 4.9e-84 490 1 BM630212 h880a03.y1 NCI CGAP
gb_est2:BG714314 + 806.00 1670.03 8.9e-84 754 1 BG714314 602664005F1 NIH_MGC_96
gb_est2:BG776825 + 805.00 1668.13 1.1e-83 741 1 BG776825 602664005F1 NIH_MGC_59
gb_est2:BF697317 + 799.00 1653.87 7.1e-83 866 1 BF697317 602129857F1 NIH_MGC_56
gb_est2:BG433851 + 797.00 1650.09 1.2e-82 1010 1 BG433851 AV7049375F1 NIH_MGC_7
gb_est1:AV714563 + 797.00 1654.15 6.9e-83 576 1 AV714563 AV714563 DCB Homo sap
```

```
gb_est2:BG494271 + 796.00 1647.91 1.5e-82 841 1 BG494271 602539211F1 NIH_MGC
gb_est2:BG030517 + 793.00 1640.56 3.9e-82 927 1 BG030517 602298055F1 NIH_MGC
gb_est2:BG708645 + 792.00 1642.50 3.1e-82 642 1 BG708645 602672568F1 NIH_MGC
gb_est2:BG564114 + 792.00 1641.93 3.3e-82 676 1 BG564114 602586171F1 NIH_MGC
gb_est2:BI596496 + 792.00 1640.85 3.8e-82 746 1 BI596496 60324353F1 NIH_MGC
```

seq_name: gb_est2:BG703377

seq_documentation_block:

```
LOCUS      BG703377          587 bp      mRNA      linear      EST 07-MAY-2001
DEFINITION 602685165F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817561 5',
ACCESSION  BG703377
VERSION    BG703377.1 GI:13975647
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
```

REFERENCE

1 (bases 1 to 587)
NIH-MGC Institute of Health, Mammalian Gene Collection (MGC)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapds@mail.nih.gov

AUTHORS

NIH-MGC Institute of Health, Mammalian Gene Collection (MGC)
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)

JOURNAL

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM0718 row: a column: 18

High quality sequence stop: 587.
location/Qualifiers
1. 587

FEATURES

source

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone_image="IMAGE:4817561"  
/clone_lib="NIH_MGC_95"  
/tissue_type="hippocampus"  
/lab_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.5 kb and  
normalized to 10^5. This is a primary library enriched  
for full-length clones and constructed using the  
cap-trapped method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH_MGC Library."  
BASE COUNT 156 a 108 c 157 g 166 t  
ORIGIN
```

alignment_scores:

Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718A-1 x BG703377 ..
Align seg 1/1 to: BG703377 from: 1 to: 587

```
1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17  
|||||  
30 ATGCGCGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 79  
|||||  
17 uUeUeUaUaUaUaUaUaUaUaUaUaUaUaUaUaUaUaUaUaUaUa 34  
|||||  
80 GTGTTGGCGGAGCTGCTTCAAGCGGTGCTTGGCGGAGGAGGAGGAGG 129
```

```

34  erSerGluAlaGlyCysArgGluLeuGlyPheSerSerSerLeuLeuLeuCysSer 50
    |||
130  CACGAGAGCATGTCAGAGAGATTAGGCTTTCTAGCACTGCTTGCAGC 179
    |||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCys 67
    |||
180  TCTTGTGATCTTCTCGGACAGTTCACCTGCTTACCTGATCTGATG 229
    |||
67  sArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
    |||
230  CAGAGAGATGCTGTCAGAGAGAACCAATTTGAACCAAAAAGCTGATG 279
    |||
84  lAglyAlaIleLeuGluValCysGly***LysLeuGlyArgPheProGln 100
    |||
280  CAGGAGCTATCTTGAAGTTGTGATGAATAATGGAAAGTTCCCTCAA 329
    |||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
    |||
330  GTCCAGAGCTTTGTTAGAGATGATTAACCAACTGTTGAGAGACTGCA 379
    |||
117  nIleLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
    |||
380  AATCAAGTATGTCCTGTTGACAGCCCTGATTAAGCTTTTGACGACA 429
    |||
134  sngLysnIleAlaGluGluLeuSerLileuLysTrpAsnTrpAspSer 150
    |||
430  ATGGGAACATGCTGTAAGACTGAGCATCTCAATGAAGAACACAGACAGT 479
    |||
151  ValGluGluPheLeuSerGluLysLeuGluArgIle 162
    |||
480  GTAGAGAAATCTCTGAGTGAATAAGTTGGAACGCATA 515
    |||
seq_name: gb_est2:BG399386

seq_documentation_block:
LOCUS      BG399386                642 bp      mRNA      linear      EST 12-MAR-2001
DEFINITION 602411143P1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4556915 5',
            mRNA sequence.
ACCESSION  BG399386
VERSION    BG399386.1 GI:13292834
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE     NIH-MGC http://nigc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cga@bbs-remail.nih.gov
          Tissue Procurement: CLOnTECH Laboratories, Inc.
          CDNA Library Preparation: CLOnTECH Laboratories, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LUCM1260 row: e column: 12
          High quality sequence stop: 634.

FEATURES
SOURCE
    location/Qualifiers
    1..642
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4556915"
    /clone_id="NIH_MGC_75"
    /lab_host="DH10B (TL phage-resistant)"
    /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
    SfiI (ggcgccctggccc); Site_2: SfiI (ggccattatggcc); 5' and
    3' adaptors were used in cloning as follows: 5' adaptor
    sequence: 5'-CACGGCATATATGGCC-3' and 3' adaptor sequence:
    5'-ATTCATAGAGCGCCAGAGCGCGCCGACATG-dT(30)BN-3' (where B = A,

```

```

C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT      166 a      116 c      165 g      195 t
ORIGIN

alignment_scores:
    Quality:      819.00      Length:      162
    Ratio:        5.119
    Percent Similarity: 98.765      Percent Identity: 98.765

alignment_block:
US-09-676-718a-1 x BG399386

Align seg 1/1 to: BG399386 from: 1 to: 642

1  MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
    |||
15  ATGCGGCTGCGCCGAGTGGGTGCTGTGTCGCGCTTGGCTACGGTT 64
    |||
17  uLeuAlaIthrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
    |||
65  GTTGTGGCGACGTGCTCTCAAGCGGTGCTGCTTTGGGCGACGCTTTT 114
    |||
34  erSerGluAlaGlyCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
    |||
115  CACGAGAGCATGTCAGAGAGAACCAATTTGAACCAAAAAGCTGATG 164
    |||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCys 67
    |||
165  TCTTGTGATCTTCTCTCGACAGTTCACCTGCTTACCTGATCTGATG 214
    |||
67  sArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
    |||
215  CAGAGAGATGCTGTCAGAGAGAACCAATTTGAACCAAAAAGCTGATG 264
    |||
84  lAglyAlaIleLeuGluValCysGly***LysLeuGlyArgPheProGln 100
    |||
265  CAGGAGCTATCTTGAAGTTGTGATGAATAATGGAAAGTTCCCTCAA 314
    |||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
    |||
315  GTCCAGAGCTTTGTTAGAGATGATTAACCAACTGTTGAGAGACTGCA 364
    |||
117  nIleLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
    |||
365  AATCAAGTATGTCCTGTTGACAGCCCTGATTAAGCTTTTGACGACA 414
    |||
134  sngLysnIleAlaGluGluLeuSerLileuLysTrpAsnTrpAspSer 150
    |||
415  ATGGGAACATGCTGTAAGACTGAGCATCTCAATGAAGAACACAGACAGT 464
    |||
151  ValGluGluPheLeuSerGluLysLeuGluArgIle 162
    |||
465  GTAGAGAAATCTCTGAGTGAATAAGTTGGAACGCATA 500
    |||
seq_name: gb_est2:BG779736

seq_documentation_block:
LOCUS      BG779736                664 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602668566F1 NIH_MGC_60 Homo sapiens CDNA clone IMAGE:4808032 5',
            mRNA sequence.
ACCESSION  BG779736
VERSION    BG779736.1 GI:14050053
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```


AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-rl@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: BLCM1660 row: d column: 17
 High quality sequence stop: 660.
 Location/Qualifiers

FEATURES

source

1..664

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4808032"

/clone_lib="NIH-MGC_60"

/tissue_type="adenocarcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site:1: SfiI (ggccgctcgcc); Site:2: SfiI (ggccattggc)

1; Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATCTAGAGCGCGAGCGCGAGCTG(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 179 a 118 c 170 g 197 t

ORIGIN

alignment_scores: Quality: 819.00 Length: 162 Ratio: 5.119 Gaps: 0 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block: US-09-676-718A-1 x BG779736 ..

Align seg 1/1 to: BG779736 from: 1 to: 664

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17
|||||
26 ATGGCGGCTGGCGGAGTGGTGTCTGTGCTCCGCGCTTTGGCGTACGCTT 75
|||||
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
|||||
76 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGACAGTTT 125
|||||
34 eSerGlnAlaCysArgGlnLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
126 CARGCGAGCATGCAGAGAGTAGGCTTTCTAGCACTTGCTTGCAGC 175
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
|||||
176 TCTTGTGATCTTCTCGACAGTTCACCTCTCAGCTGATCTGATTG 225
|||||
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGluThrLysLysLeuYrA 84
|||||
226 CAGAGCATGCTGTACAGAGAGACACAAATTGAAACCAAAAGCTGATG 275
|||||
84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
|||||
276 CAGGACATATCTTGAAGTTGTGATGAATAATTGGAGAGTTCCTCAA 325
|||||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117

```

seq_name: gb_est1:AV757896

seq_documentation_block:

LOCUS AV757896 694 bp mRNA linear EST 19-OCT-2000

DEFINITION AV757896 BM Homo sapiens cDNA clone BMFALD04 5', mRNA sequence.

ACCESSION AV757896

VERSION AV757896.1 GI:10915744

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 694)

AUTHORS Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H., Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gao,G., Wang,Z., Zhang,Q., Chen,S., Han,Z. and Chen,Z.

Human sapiens cDNA BM clones

Unpublished (2000)

CONTACT: Zengqiang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzq@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..694

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="BMFALD04"

/clone_lib="BM"

/tissue_type="Bone marrow"

/cell_type="CD34+ hematopoietic stem/progenitor cell"

/lab_host="BM25.8"

/note="Vector: pTriplEx2; Site:1: SfiI; Site:2: SfiI"

BASE COUNT 163 a 139 c 189 g 201 t

ORIGIN

alignment_scores: Quality: 819.00 Length: 162 Ratio: 5.119 Gaps: 0 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block: US-09-676-718A-1 x AV757896 ..

Align seg 1/1 to: AV757896 from: 1 to: 694

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17
|||||
105 ATGGCGGCTGGCGGAGTGGTGTCTGTGCTCCGCGCTTTGGCGTACGCTT 154
|||||
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
|||||
155 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGACAGTTT 204
|||||

```

34 erserGluaIaCysArGluLeuGlyPheSerSerAsnLeuLeuCySer 50
|||||
205 CATCGAGGATGATGAGAGAGATTAGCTTTCTAGCAACTGCTTGCACG 254
51 SerCysAspLeuLeuGlyInPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
255 TCTTGATCTCTCGGACAGTTCAACCTGCTTCAGCTGCATCTGATG 304
67 sArGglCyScyGInGluGluAlaGlnPheGluThrLysLysLeuTyra 84
305 CAGAGAGATGCTGTCAGAGAGACCAATTGAAACCAAAAAGCTGTATG 354
84 lAgIAlaIleLeuGluValCySGly**LysLeuGlyArGpPheProGln 100
355 CAGGAGCTATTCTTGAGATTGTGGATGAATAATGGGAAGGTTCCCTCA 404
101 ValGlnAlaPheValArGSerAspLysProLysLeuPheArGlyLeuGln 117
405 GTCCAAAGCTTTTGTAGAGATGATTAACCAACCAAGCTTCAGAGACTGCA 454
117 nIleLysTyValArGlySerAspProValLeuLysLeuLeuAspAspA 134
455 AATCAAGATATGTCGCGTTTCAGACCCCTGTATTAAAGCTTTTGACGACA 504
134 sncGLyAsnIleAlaGluGluLeuSerLleLeuLysTyPAsnThrAspSer 150
505 ATGGAACTTCTCTGAGAGACTGAGCATTTCTCAATGGAAACAGACAGCT 554
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
555 GTAGAAAGATTCCTGAGTGAAAGTTGGAAGCAGATA 590
seq_name: gb_est1:AV718270
seq_documentation_block:
LOCUS AV718270 700 bp mRNA linear EST 16-OCT-2000
DEFINITION AV718270 FHTB Homo sapiens cDNA clone FHTBABF07 5', mRNA sequence.
ACCESSION AV718270
VERSION AV718270.1 GI:10815422
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 700)
AUTHORS Xiao,H., Peng,Y., Song,H., Gu,Y., Yang,Y., Gao,G., Xu,X., Li,N.,
Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Chen,Z. and Han
,Z.
TITLE Homo sapiens cDNA FHTB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
331 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@cnhg.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..700
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FHTBABF07"
/clone_id="FHTB"
/tissue_type="hypothalamus"
/dev_stage="Fetal"
/lab_host="BM25.8"
/note="Vector: pTRIPLEX2; Site_1: sfilA; Site_2: sfilB1"
BASE COUNT 191 a 130 c 171 g 204 t 4 others
ORIGIN

alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765
alignment_block:
US-09-676-718a-1 x AV718270 ..
Align seg 1/1 to: AV718270 from: 1 to: 700
1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
14 ATGGCGGCTGGGCGGAGTGGGTCTGTCTGTCGGCGCTTTGGGCTACGGTT 63
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
64 GTTGTTGGGAGCTGTGCTTCAGCGGCTGTCTGCTTTGGGCGAGAGTTT 113
34 erserGluaIaCysArGluLeuGlyPheSerSerAsnLeuLeuCySer 50
114 CATCGAGGATGATGAGAGAGATTAGCTTTCTAGCAACTGCTTGCACG 163
51 SerCysAspLeuLeuGlyInPheAsnLeuLeuGlnLeuAspProAspCy 67
164 TCTTGATCTCTCGGACAGTTCAACCTGCTTCAGCTGCATCTGATG 213
67 sArGglCyScyGInGluGluAlaGlnPheGluThrLysLysLeuTyra 84
214 CAGAGAGATGCTGTCAGAGAGACCAATTGAAACCAAAAAGCTGTATG 263
84 lAgIAlaIleLeuGluValCySGly**LysLeuGlyArGpPheProGln 100
264 CAGGAGCTATTCTTGAGATTGTGGATGAATAATGGGAAGGTTCCCTCA 313
101 ValGlnAlaPheValArGSerAspLysProLysLeuPheArGlyLeuGln 117
314 GTCCAAAGCTTTTGTAGAGATGATTAACCAACCAAGCTTCAGAGACTGCA 363
117 nIleLysTyValArGlySerAspProValLeuLysLeuLeuAspAspA 134
364 AATCAAGATATGTCGCGTTTCAGACCCCTGTATTAAAGCTTTTGACGACA 413
134 sncGLyAsnIleAlaGluGluLeuSerLleLeuLysTyPAsnThrAspSer 150
414 ATGGAACTTCTCTGAGAGACTGAGCATTTCTCAATGGAAACAGACAGCT 463
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
464 GTAGAAAGATTCCTGAGTGAAAGTTGGAAGCAGATA 499
seq_name: gb_est2:BI518529
seq_documentation_block:
LOCUS BI518529 726 bp mRNA linear EST 29-AUG-2001
DEFINITION 603061650F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5210821 5',
mRNA sequence.
ACCESSION BI518529
VERSION BI518529.1 GI:15343321
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 726)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM11529 row: c column: 14
 High quality sequence stop: 724.

FEATURES

SOURCE

Location/Qualifiers
 1..726
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5210821"
 /clone_lib="NIH_MGC_95"
 /tissue_type="leukocyte"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
 (destroyed); RNA source: leukocytes from anonymous pool of
 non-activated adult donors. Library is oligo-dT primed
 and directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 1.2-3.3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 027. Note:
 this is a NIH_MGC Library."
 BASE COUNT 196 a 134 c 172 g 224 t
 ORIGIN

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x B1518529 ..

Align seg 1/1 to: B1518529 from: 1 to: 726

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||
2 ATGGCGGCTGGGCGGAGTGGTCTGTGTCGCCGCTTGGCTACGGTT 51
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
  |||
52 GTTGTGGCGACTGTCCTTCACAGCGGTCTGCTTTGGGCGAGAGTTT 101
34 erserGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
  |||
102 CATTGGAGGCGATGCAGAGATTAGGCTTTCTAGCAACTTCTTGCAGC 151
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCys 67
  |||
152 TCTTGTGATCTTCTCGAGAGTTCACACTGCTTCACTGATCTGATTG 201
67 sArgGlyCysGlnGlnGluAlaGlnPheGluThrLysLysLeuTyr 84
  |||
202 CAGAGGATGCTGTCAGGAGGACACAAATTGAACCAAAACCTGATG 251
84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
  |||
252 CAGGACCTATTCTTGAAGTTTGGATGAATAATTGGGAAGTCCCTCAA 301
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
  |||
302 GTCCAAAGCTTTGTAGAGTGAATAAACCAACTGTTCAAGAGACTGCA 351
117 nileLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAs 134
  |||
352 AATCAAGTATGTCGTGGTTCAGACCCGTATTAAGCTTTTGGAGCACA 401
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrPheAsnThrAs 150
  |||
402 ATGGGAACAATTTGCTGAAGAAGTACGATTTCTCAATGGAACACAGAC 451
  |||

```

151 ValGluGluPheLeuSerGlyLysLeuGluArgIle 162
 |||
 452 GTAGAAGAAATCTCGAGTGAATAAGTTGGAACGCATA 487

seq_name: gb_est2:B1545979

seq_documentation_block:

LOCUS B1545979 739 bp mRNA linear EST 05-SEP-2001
 DEFINITION 603188102P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259576 5',
 mRNA sequence.
 ACCESSION B1545979
 VERSION B1545979.1 GI:15433291
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 739)

AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bcrfmail.nih.gov
 Tissue procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA library preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshitsuki and Piero Carninci (RIKEN)
 cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM11654 row: c column: 01
 High quality sequence stop: 737.

FEATURES

SOURCE

Location/Qualifiers
 1..739

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5259576"
 /clone_lib="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-gtgttttttttttttyn-3',
 size-selected for average insert size 2.5 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 198 a 135 c 178 g 228 t
 ORIGIN

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x B1545979 ..

Align seg 1/1 to: B1545979 from: 1 to: 739

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||
19 ATGGCGGCTGGGCGGAGTGGTCTGTGTCGCCGCTTGGCTACGGTT 68
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
  |||
69 GTTGTGGCGACTGTCCTTCACAGCGGTCTGCTTTGGGCGAGAGTTT 118
34 erserGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
  |||

```

119 CATGGAGGATGCGAGAGTAGTGGCTTTCTGTGCAACTGCTTGGAC 168
116 SerCysAspLeuLeuGlnGlnPheAsnLeuGlnLeuAspProAspCys 67
115 TCTTGTATCTCTCGGACAGTTCAACTCTCTCAGCTGGATCTGTATTG 218
169 sArgGlyCysCysGlnGlnGlnIuaIaGlnPheGlnuThrLysLysLeuTyrA 84
219 CAGAGCATGCTGTGCAGAGGAGACACAAATTGGAACCAAAAACCTGTATG 268
84 IacGlyAlaIleLeuGlnIuaLysCysGly**LysLeuGlnYArgPheProGln 100
269 CAGAGCATTCTTGAAATTGTGGATGAATAATGGGAGAGTCCCTCAA 318
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
319 GTCCAAACCTTTTGTTAGAGTAGATAAACCCAAACTGTTCAGAGAGCTGCA 368
117 nIleLysTyrValArgLysSerAspProValIleuLysLeuLeuAspAspA 134
369 AATCAAGATATGTCGCGTGCTTCAGACCCCTGTATTAAAGCTTTTGGACACA 418
134 snGlyAsnIleAlaGlnGlnLeuSerIleLeuLysTrpAsnThrAspSer 150
419 ATGGAGCAATCTGCTGAAGAACTGAGCTTCTCAATATGAAACACAGACAGT 468
151 ValGlnIuPheLeuSerGluLysLeuGlnuThrIle 162
469 GTACAGACAAATTCCTGAGTAGAAAGTTGGAAAGGCATTA 504

name: gb_est2.BG505091

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4664195"
/clone_lib="NH_MGC_61"
/lssue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site.1: SfiI (ggcgccctggcgc); Site.2: SfiI (ggcctctggcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dt(30)BN-3',

```

(where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MCC Library."

```
alignment_scores:
  Quality: 819.00
  Ratio: 5.119
  Gaps: 0
Percent Similarity: 98.765
Percent Identity: 98.765
```

alignment_block:
US-09-676-718A-1 x BG505091

Align seg 1/1 to: BG505091 from: 1 to: 743

1 MetAlaIaAGIProSerClyCysLeuValProAlaPheGlyLysArgLe 17
161 AATGAGGCGCTGGCCGACGTGGTGTGTGTGTCGGCGGCTTGGCGTACCGTT 111
15 ATGGCGGCGCTGGCCGACGTGGTGTGTGTGTCGGCGGCTTGGCGTACCGTT 64
17 uLeuLeuAlaThrValIleuGlnAlaValSerAlaPheGlyAlaIuPheS 34
65 GTTGTGGGAGCTGTGCTTCAAGCGGTGTGCTTGGGGGACAGATTTT 114
34 eTserLualAcysArgIuIeuglyPheSerSerAsnLeuLeuCysSer 50
115 CACGGAGGCAATGCAGAGAGTTCAGCTTTTCACAGCACTTCCTTGCAGC 164
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
165 TCTTGTGATCTTCTCGGACAGCTTCAACTGCTTCACTGGAATCTGTGATG 214
67 sArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyra 84
215 CAGAGATGCTGTGCAGAGAGGAGCAATTTGAACCAAAAGCTGTATG 264
84 IaGlyAlaIleIeuGlnValCysGly***LysLeuGlyValArgPheProGln 100
265 CAGAGACTTCTTGTGAAGTTGTGATGAAATATGGGAAGTCTCCCTCAA 314
315 GTCCAAAGCTTTGTTAGGAGTATATAACCCAAACGTTCACAGAGCATGCA 364
117 nileLysTyraValArgGlySerAspProValIleuLysLeuLeuAspAspA 134
365 AATCAGCATATGTCGCGGTTCAGACCCGTATTAAGCTTTTGGACACACA 414
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTTPAsnThrAspSer 150
415 ATGGAGCACTTGTCTGAAGAACTAGACATCTTCCAATGGAACACAGACAGT 464
151 ValGluGluPheLeuSerGlnLysLeuAluArgTle 162
465 GTTGAAGCAATTCCTGATGTAATAAGTTGAAGCGCATA 500

seq_name: gb_est2:BG613877

seq_documentation_block: 744 bp mRNA linear EST_18-APR-2001

LOCUS BG613877

DEFINITION 60263968AF1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4770860 5',

ACCESSION BG613877

VERSION BG613877.1 GI:13665248

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 744)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Plate: LNCM1638 row: g column: 21
 High quality sequence stop: 714.
 Location/Qualifiers
 1..744

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4770860"
 /clone_lib="NIH_MGC_61"
 /tissue_type="embryonal carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccctcgcc); Site_2: SfiI (ggcattatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library."
 BASE COUNT 204 a 132 c 182 g 226 t
 ORIGIN

alignment_scores:
 Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718A-1 x BG613877 ..

Align seg 1/1 to: BG613877 from: 1 to: 744

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyIysArgLe 17
|||||
27 ATGGCGGCTGGCCGAGTGGTGTCTGTCCGCGCTTGGGCTACGGTT 76
|||||
17 ulcLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
|||||
77 GTTGTGGCCACGTGTCTCAAGCGGTGTCTGCTTTGGGGCAGAGTTT 126
|||||
34 eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
127 CATCGGAGCATGCAGAGACTTAAGCTTTCTACCACTTCCTTTCACGC 176
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
|||||
177 TCTTGTGATCTTCTCGACAGATTCACCTGCTTCAGCTGATGATG 226
|||||
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGluThrIlysIysLeuYr 84
|||||
227 CAGAGGATGCTGTACAGGAGACACAAATTGAACCAAAAAGCTGATG 276
|||||
84 laGlyAlaIleLeuGluValCysGly**IysLeuGlyArgPheProGln 100
|||||
277 CAGAGCATTTCTTGAAGTTTGTGATGAAGAAATTGGGAGAGTCCCA 326

```

```

101 ValGlnAlaPheValArgSerAspIysProIysLeuPheArgGlyLeu 117
|||||
327 GTCCACACTTTTGTAGAGGATAAACCAACTGTTCAGAGCACTTCA 376
|||||
117 nIleIysTyValArgGlySerAspProValLeuIysLeuAspAspA 134
|||||
377 AATCAAGTATGCTCCGTGCTCAGACCCGTATTAAGCTTTGGACACA 426
|||||
134 snGlyAsnIleAlaGluLeuSerIleLeuIysTrpAsnThrAspSer 150
|||||
427 ATGGGAACATGCTGAGAACTGAGCATTCCTCAATGAAACACAGACAGT 476
|||||
151 ValGluGluPheLeuSerGluIysLeuGluArgIle 162
|||||
477 GTAGAGAAATTCCTGAGTGAAGTTGGACGACATA 512
|||||
seq_name: gb_est2:BG777365

```

seq_documentation_block: 746 bp mRNA linear EST 15-MAY-2001
 LOCUS BG777365
 DEFINITION 602664622F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4804732 5', mRNA sequence.
 ACCESSION BG777365
 VERSION BG777365.1 GI:14047682
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 746)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DP
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Plate: LNCM1651 row: k column: 05
 High quality sequence stop: 731.
 Location/Qualifiers
 1..746

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4804732"
 /clone_lib="NIH_MGC_60"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccctcgcc); Site_2: SfiI (ggcattatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC library."
 BASE COUNT 203 a 139 c 180 g 224 t
 ORIGIN

alignment_scores:
 Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:
US-09-676-718a-1 x BG777365 ..

Align seg 1/1 to: BG777365 from: 1 to: 746

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||||||
26 ATGGCGGCGTGGCCGAGTGGTGTCTGTCGCGCGCTTGGCGCTACGGTT 75
  |||||||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
  |||||||
76 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGACTTTT 125
  |||||||
34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer 50
  |||||||
126 CATCGAGGCGATGACAGAGATTAGGCTTTCTAGCAACTGCTTTCACGC 175
  |||||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProGly 67
  |||||||
176 TCTTGTGATCTTCTCGACAGATTCAACCTGCTTCACTGATCTGATTTG 225
  |||||||
67 sArgGlyCysCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
  |||||||
226 CAGAGGATGCTGTCAAGAGAGACCAATTGAAACCAAAAGCTGTATG 275
  |||||||
84 laglAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
  |||||||
276 CAGGAGCTATTCTTGAAAGTTGTGATGAAATGGGAAGGTTCCCTCAA 325
  |||||||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
  |||||||
326 GTCCAAAGCTTTGTTAGAGAGTATAAACCAACTGTTCAAGAGACTGCA 375
  |||||||
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 134
  |||||||
376 AATCAAGATATGTCGGTGTACAGCCCTGTATTAAGCTTTTGGACGACA 425
  |||||||
134 sngLysAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
  |||||||
426 ATGGGAACATTGCTGAAAGATGAGCATTTCTCAAAATGGAAACAGACAG 475
  |||||||
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
  |||||||
476 GTAGAAGAAATTCCTGAGTCAAAAGTTGGAACGCAAT 511
  |||||||
seq_name: gb_est2:BI597212
seq_documentation_block:
LOCUS BI597212 750 bp mRNA linear EST 07-SEP-2001
DEFINITION 603250980F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5302891 5',
mRNA sequence.
ACCESSION BI597212
VERSION BI597212.1 GI:15490151
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apds-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NIGRI), Shiraki
Toshuyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
cDNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHAM1766 row: 0 column: 20

```

High quality sequence stop: 747.
FEATURES
source Location/Qualifiers
1..750

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5302891"
/clone_id="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptPR (modified pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to R0.5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIGRI/NIH). National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 202 a 140 c 183 g 225 t
ORIGIN

alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:
US-09-676-718a-1 x BI597212 ..

Align seg 1/1 to: BI597212 from: 1 to: 750

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||||||
34 ATGGCGGCGTGGCCGAGTGGTGTCTGTCGCGCGCTTGGCGCTACGGTT 83
  |||||||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
  |||||||
84 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGACTTTT 133
  |||||||
34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer 50
  |||||||
134 CATCGAGGCGATGACAGAGATTAGGCTTTCTAGCAACTGCTTTCACGC 183
  |||||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProGly 67
  |||||||
184 TCTTGTGATCTTCTCGACAGATTCAACCTGCTTCACTGATCTGATTTG 233
  |||||||
67 sArgGlyCysCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyr 84
  |||||||
234 CAGAGATGCTGTCAAGAGAGACCAATTGAAACCAAAAGCTGTATG 283
  |||||||
84 laglAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
  |||||||
284 CAGGAGCTATTCTTCAAGTTTGTGATGAAATTTGGGAAGGTTCCCTCAA 333
  |||||||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
  |||||||
334 GTCCAAAGCTTTGTTAGAGAGTATTAACCAACTGTTCAAGAGACTGCA 383
  |||||||
117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 134
  |||||||
384 AATCAAGATATGTCGGTGTACAGCCCTGTATTAAGCTTTTGGACGACA 433
  |||||||
134 sngLysAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150
  |||||||
434 ATGGGAACATTGCTGAAAGATGAGCATTTCTCAAAATGGAAACAGACAG 483
  |||||||
151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
  |||||||
484 GTAGAAGAAATTCCTGAGTCAAAAGTTGGAACGCAAT 519
  |||||||
seq_name: gb_est1:AL533548

```

```

seq.documentat_block:
LOCUS      AL533548              787 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL533548 LTI_FL015_Brn1 Homo sapiens cDNA clone CS0DN004YB08 5
prime, mRNA sequence.
ACCESSION  AL533548
VERSION     AL533548.1  GI:12797041
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 787)
AUTHORS     Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 Evry cedex - France
             Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

FEATURES
source
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="LTI_FL015_Brn1"
/sex="male"
/tissue_type="Adult brain"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was constructed by Life Technologies.
Contact : Feng Liang Life Technologies, a division of
Invitrogen 9800 Medical Center Drive Rockville, Maryland
20850, USA Fax : (1) 301 610 8371 Email :
fliang@litech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT      216 a      140 c      185 g      246 t

alignment_scores:
Quality:      819.00      Length:      162
Ratio:        5.119      Gaps:      0
Percent Similarity: 98.765      Percent Identity: 98.765

alignment_block:
US-09-676-718a-1 x AL533548 ..

Align seg 1/1 to: AL533548 from: 1 to: 787
1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
|||||
6 ATGGCGGCTGGGCGGAGTGGGTCTGTGGTCCGGCGCTTTGGGCTACGGTT 55
|||||
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
|||||
56 GTTGTGGGAGCTGTGCTTCAAGCGGTGCTGCTTTGGGGCAGAGTTT 105
|||||
34 eSerGlnAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
106 CATCGAGGACATGCAGAGAGTAGGCTTTTCAGCAACTGCTTTCACAG 155
|||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
156 TCTTGTGATCTTCTCGGACAGTTCAACCTGCTTACACTGATCTCATTTG 205
|||||
67 sArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyrA 84
|||||
206 CAGAGATGCTGTCAGAGAGAACAAATTTGAAACCAAAAAGCTGATG 255
|||||
84 lglYAlaIleuGlnuValCysGly**LysLeuGlyArgPheProGln 100

```

```

seq.name: gb_est2:BG502649

seq.documentat_block:
LOCUS      BG502649              797 bp      mRNA      linear      EST 27-MAR-2001
DEFINITION 602549479P1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4657338 5',
mRNA sequence.
ACCESSION  BG502649
VERSION     BG502649.1  GI:13464166
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 797)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-remail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: IMCMI449 row: m column: 19
             High quality sequence stop: 753.

FEATURES
source
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:4657338"
/clone_id="NIH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (TI Phage-resistant)"
/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccctcgcc); Site_2: SfiI (ggccatlaagcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTTAGAGCGCGCGCGCATG-3' and 3' adaptor
sequence: 5'-ATCTTAGAGCGCGCGCGCATG-3' (30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
library."

BASE COUNT      218 a      141 c      192 g      246 t

ORIGIN
1
|||||
alignment_scores:
Quality:      819.00      Length:      162

```

Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765
 alignment_block:
 US-09-676-718A-1 x BG502649 ..

Align seg 1/1 to: BG502649 from: 1 to: 797

```

1  MetaAlaIaglyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
   |||||
15  ATGCGCGCTGGCGCGAGTGGGTCTGTCGCCGCGTTGGCGTACGGTT 64
   |||||
17  uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyLagLuphes 34
   |||||
65  GTTGTGGCGACTGCTCTCAAGCGGTCTGCTTTGGGGCAGAGTTT 114
   |||||
34  erSerGluAlaCysArgGluLeuGlyLysPheSerSerAsnLeuLeuCysSer 50
   |||||
115  CATCGAGGCGATGCGAGAGATTAGGCTTTCTAGCAACTGCTTTCGACG 164
   |||||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
   |||||
165  TCTTGTGATCTTCTCGGACAGTCAACCTGCTCAGCTGATCCGATTG 214
   |||||
67  sArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyrA 84
   |||||
215  CAGAGGATGCTGTCAGAGAGACCAATTGAAACCAAAAGCTGTAATG 264
   |||||
84  laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
   |||||
265  CAGAGAGCTATTCTTGAAGTTGTGGATGAAATGGGAAGTTCCCTCAA 314
   |||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
   |||||
315  GTCCAAAGCTTTGTAGAGATGAATAAACCAAGCTTCAGAGAGCTGCA 364
   |||||
117  nileLysTyrValArgGlySerAspProValLeuLeuLeuAspAspA 134
   |||||
365  AATCAAGATATCTCGGTGTCAGACCTGTAATTAAGCTTTGGACGACA 414
   |||||
134  snGlyAsnIleAlaGluLuleuSerIleLeuLysTyrAsnThrAspSer 150
   |||||
415  ATGGGAACATTGCTGAAGAACTGAGCATTCCTCAATGGAACACAGACAGT 464
   |||||
151  ValGluGluPheLeuSerGluLysLeuGluArgTle 162
   |||||
465  GTAGAGAAATTCCTGAGTGAAGTGGAAACGACATCA 500
   |||||

```

seq_name: gb_est2:BG533149

seq_documentation_block:

LOCUS BG533149 802 bp mRNA linear EST 03-APR-2001
 DEFINITION 60280673F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4718490 5',
 mRNA sequence.

ACCESSION BG533149
 VERSION BG533149.1 GI:13524689
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 802)
 AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: CLOVETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incycle Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>
 Plate: LDCM1570 row: a column: 19
 High quality sequence stop: 723.
 Location/Qualifiers
 1..802

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4718490"
 /clone_1ib="NIH_MGC_61"
 /tissue_type="embryonal carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: testis; Vector: pMDR-11B (Clontech); Site_1:
 SfiI (ggccgctggcc); Site_2: SfiI (ggccattggcc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CAGCGCCATATGCGC-3' and 3' adaptor
 sequence: 5'-ATTTCAGAGCGCGGCGGCGGCACATG-dT(30)-BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library."

BASE COUNT 222 a 142 c 197 g 241 t
 ORIGIN

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718A-1 x BG533149 ..

Align seg 1/1 to: BG533149 from: 1 to: 802

```

1  MetaAlaIaglyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
   |||||
34  ATGCGCGCTGGCGCGAGTGGGTCTGTCGCCGCGCTTGGCGTACGGTT 83
   |||||
17  uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyLagLuphes 34
   |||||
84  GTTGTGGCGACTGCTCTCAAGCGGTCTGCTTTGGGGCAGAGTTT 133
   |||||
34  erSerGluAlaCysArgGluLeuGlyLysPheSerSerAsnLeuLeuCysSer 50
   |||||
134  CATCGAGGCGATGCGAGAGATTAGGCTTTCTAGCAACTGCTTTCGACG 183
   |||||
51  SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
   |||||
184  TCTTGTGATCTTCTCGGACAGTCAACCTGCTCAGCTGATCCGATTG 233
   |||||
67  sArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLeuTyrA 84
   |||||
234  CAGAGGATGCTGTCAGAGAGACCAATTGAAACCAAAAGCTGTAATG 283
   |||||
84  laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
   |||||
284  CAGAGAGCTATTCTTGAAGTTGTGGATGAAATGGGAAGTTCCCTCAA 333
   |||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
   |||||
334  GTCCAAAGCTTTGTAGAGATGAATAAACCAAGCTTCAGAGAGCTGCA 383
   |||||
117  nileLysTyrValArgGlySerAspProValLeuLeuLeuAspAspA 134
   |||||
384  AATCAAGATATCTCGGTGTCAGACCTGTAATTAAGCTTTGGACGACA 433
   |||||
134  snGlyAsnIleAlaGluLuleuSerIleLeuLysTyrAsnThrAspSer 150
   |||||
434  ATGGGAACATTGCTGAAGAACTGAGCATTCCTCAATGGAACACAGACAGT 483
   |||||

```


151 ValGlugluPheIeuSerGluLysLeuGluArgIle 162
 |||
 484 GTAGAAAGATTCTCGAGTGAAGTGAACGACATA 519

seq_name: gb_est2:BG529329

seq_documentation_block:

LOCUS BG529329 808 bp mRNA linear EST 03-APR-2001
 DEFINITION 602558441F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4687019 5',
 mRNA sequence.

ACCESSION BG529329

VERSION BG529329.1 GI:13520866

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 808)

NIH-MGC http://mgc.ncl.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.
 Email: cgaabp-r@mail.nih.gov

TISSUE Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LHCMI499 row: b column: 12

High quality sequence stop: 732.

Location/Qualifiers

1..808

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image:4687019"

/clone_lib="NIH_MGC_61"

/tissue_type="embryonal carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:

SfiI (ggcgatcggcc); Site_2: SfiI (ggcattatggcc);

double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-ATTCGAGCGCCGACGACATG-dT(30)BN-3'

(where B = A, C, G, or T). Average

insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH-MGC

library."

BASE COUNT 222 a 145 c 200 g 241 t

ORIGIN

alignment_scores: Quality: 819.00 Length: 162

Ratio: 5.119 Gaps: 0

Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x BG529329 ..

Align seg 1/1 to: BG529329 from: 1 to: 808

1 MetAlaAlaIyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17

|||||

16 ATGGCGCGCTGGCGGACGCTGCTGCTGCGCGCTTTGGGCTACGCTT 65

|||||

17 uLeuIeuAlaThrValIeuGlnAlaIserAlaPheGlyAlaIuphes 34

|||||

66 GTTGTTGGCAGACTGCTTCAAGCGGTGCTGCTTTTGGGCGACAGATT 115

34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
 |||

116 CATCGAGAGCATCGACAGAGATTAGCTTTTCTACCAACTCTCTTGCACG 165

|||||

51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCys 67

|||||

166 TCTTGTGATTCCTTCGACAGTTCACCTGCTTCAGTGGATCTGATTTG 215

|||||

67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyrA 84

|||||

216 CAGAGATGCTGCTCAGAGAGACACAAATTGAAACCAAAAAGCTGTTATG 265

|||||

84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100

|||||

266 CAGAGCATATTCCTTGAAGTTTGTGATGAAATAATTGGCAAGCTCCCTCA 315

|||||

101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117

|||||

316 GTCCAAAGCTTTGTAGAGATGATTAACCAACTGTTCAAGAGACTGCA 365

|||||

117 nIleLysTyrValArgGlySerAspProValIleuLysLeuLeuAspAsp 134

|||||

366 AATCAAGTATGTCCTGCTGCTCAGACCTGTATTAAGCTTTGGACGACA 415

|||||

134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTrpAsnThrAspSer 150

|||||

416 ATGGGAACATTCGTGAAGAACTGAGCAATTCCAATGGAACACAGACAGT 465

|||||

151 ValGlugluPheIeuSerGluLysLeuGluArgIle 162

|||||

466 GTAGAAAGATTCTCGAGTGAAGTGAACGACATA 501

THIS PAGE BLANK (USPTO)

OM of: US-09-676-718A-1 to: Issued_Patents_NA:* out_format : pfs
Date: Sep 18, 2002 5:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/cg2n2.1/USPRO.spool/US09676718/runat_17092002_143433_21323/app-query.fasta.1.222
-DB=Issued_Patents_NA -ORMT=fastap -SUPER=rml -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOOEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09676718 -ECGN1_1.45 -NCPU=6 -ICPU=3 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-676-718A-1
Query length: 162
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 43.970000

score_list:

Sequence	Strid Orig	ZScore	EScore	Len	Documentation	..	2.7e-65	578
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-385-982-439 +				552.50	1310.19		2.7e-65	578
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-103-840A-2 +				78.00	59.67		1.1e+05	4403765
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-972-927-1 +				77.50	137.31		5.85	5232
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-912-227-1 +				75.00	143.44		2.66	1849
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-972-927-4 +				74.50	130.19		14.56	5175
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-441-139-6 +				74.00	137.16		5.96	2574
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-178-477B-42 +				69.50	125.38		26.97	2781
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-441-139-1 +				69.00	124.18		31.47	2780
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-103-840A-1 +				68.00	35.50		7.4e+05	4411529
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-752-760A-1 +				67.00	89.68		2.6e+03	35081
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-603-150A-3 +				66.50	125.65		26.06	1464
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-942-686-1 +				66.50	117.63		72.92	2905
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-942-686-1 +				66.00	112.74		136.52	3979
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-180-439-1 +				66.00	112.74		136.52	3979
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-180-439-2 +				66.00	112.74		136.52	3979
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-180-439-7 +				65.50	107.81		257.02	5470
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-441-139-12 +				65.50	107.81		257.02	5470
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-09-328-111-458 +				65.00	133.26		9.83	561
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-058-489-26 +				65.00	124.74		29.28	1161
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-759-581B-3 +				65.00	120.42		51.00	1680
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-304-711-3 +				65.00	120.42		51.00	1680
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-07-922-724-7 +				65.00	116.42		85.11	2363
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-603-426A-7 +				65.00	116.42		85.11	2363
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-374-652C-1 +				65.00	116.34		85.97	2934
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-09-149-934-2 +				65.00	113.89		117.78	2934
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-333-553A-1 +				64.50	109.41		209.28	3881
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-444-792-3 +				63.50	113.73		120.25	2183
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-445-042-3 +				63.50	113.73		120.25	2183
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-459-25 +				63.50	105.62		340.30	4365
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-459-25 +				63.50	105.62		340.30	4365
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-781-893-7 +				63.50	105.62		340.30	4365
/cg2n2.6/ptodata/2/lna/5A.COMB.seq:US-08-781-893-7 +				63.50	105.62		340.30	4365
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-602-228-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-602-228-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-643-341A-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-643-341A-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-494-440B-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-494-440B-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-839-032A-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-839-032A-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-839-032A-7 +				63.00	121.39		44.99	1023
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-839-032A-7 +				63.00	121.39		44.99	1023

/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-568-459A-11 + 63.00 121.39 44.99 10
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-568-459A-11 + 63.00 121.39 44.99 10
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-487-826B-11 + 63.00 121.39 44.99 10
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-487-826B-11 + 63.00 121.39 44.99 10
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-487-826B-11 + 63.00 121.39 44.99 10
/cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-08-213-448-2 + 62.50 127.22 21.32 561

seq_name: /cg2n2.6/ptodata/2/lna/5B.COMB.seq:US-09-385-982-439

seq_documentation_block:

Sequence 439, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
NUMBER OF SEQ ID NOS: 344
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(578)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-439

alignment_scores:

Quality	552.50	Length:	136
Ratio:	4.604	Gaps:	3
Percent Similarity:	88.235	Percent Identity:	86.029

alignment_block:

US-09-676-718A-1 x US-09-385-982-439 ..

Align seg 1/1 to: US-09-385-982-439 from: 1 to: 578

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
|||||
43 ATGGCGGCTGGGCGGAGTGTCTGTGCGCGGCTGGGCTGACGCTT 92
17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyLysArgLe 34
|||||
93 GTTGTGGCGACTGTCTTCAAGCGGTCTGCTTTGGGCGAGGTTT 142
34 eTserGlnAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
143 CATCGAGAGCATGAGAGATTAGCTTTTACAGAACTGCTTGCAC 192
51 SerGysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
|||||
193 TCTGTGTATCTTCGCGACATTCACTGCTTACGTCGATCCGATTG 242
67 sArgGlyCysGlyGlnGluAlaGlnPheGluThrLysLysLeuTyr 84
|||||
243 CAGAGGATGCTGTCAAGAGGAGCAATTTGAAACCAAAAGCTGATG 292
84 lagAlaAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
|||||
293 CAGAGGATGCTGTCAAGAGGAGCAATTTGAAACCAAAAGCTGATG 342
101 ValGlnAlaPheValArgSerAspLysProLysLeuPhe...ArgGlyL 116
|||||

```

343 GTCACANCTTTGTANGAGTATAAACCAACCACTGTCACAGACTG 392
116 eugli1lelystyVal.ArgglySerAspProValLeuLysLeuLeu 132
333 CCAATTAAGTATGTNNCGTGTTCACAAACGTCGATTAAGGCTTT 442
132 PASP 133
443 NGAC 446

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

alignment_scores:
Quality: 78.00 Length: 141
Ratio: 1.000 Gaps: 6
Percent Similarity: 55.319 Percent Identity: 26.950

alignment_block:
US-09-676-718A-1 x US-09-103-840A-2 ..
Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765

12 AlapheglyLysArgLeuLeu.....LeuAl 20
503840 AGCTTCGCTCCGCGCTCATCTAGGTACTGGGGTGCAGCAATCTG 503889
20 aThValLeuGlnAlaValSerAlaPheGlyValGluPheSerSerGluA 37
503890 GGTGCTAGAGCAGCGCTGATGCGCTCAGGTACCGAGCTGACCGCTG 503939
37 lAcysArgGluLeuGlyPhe.....SerSerAsnLeuLeuGlySerSer 51
503940 CGATACGCGCGGCTCGACCGCAGCGGAGAACCGCGCTGCTC..... 503980
52 CysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCysAr 68
503981 ..GACCTGCTACCGGCTCGGCTACACACCGCTA...CCCAACACCGC 504024
68 ggLyCysGlyGlnGluValGlnPheGluThrLysLysLeuTYRAlag 85
504025 GGGGTGCGCGACGCGCGGGAAGCGGTCTGACAGCGAGTGGCGCGG 504074
85 lYAlaIle.....LeuGlnValCysGly**LysLeu 95
504075 AGGCGCTGAAACCAACTGGGTCAAGCTCGAGTATTGCCGACGAAAC 504124
96 GlyArgPheProGlnValGlnAlaPheValArgSerAspLysProLysLe 112

```

```

504125 ACCCTGTGGCTGATCGGCTGCAATTAGTCGGGCTGCAGAACAAATTG 504174
112 uPheArgGlyLeuGlnIleLysTyValArgLysSerAspProValLeu 128
504175 GGACGACGATTTGTGTCTTACCGTACACCAACCGACGACCGGCTG 504224
129 ..LysLeuLeuAspAspAsnGly 135
504225 CCCGCCGCTAGAGATACCGGT 504247

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-972-927-1

seq_documentation_block:
; Sequence 1, Application US/08972927
; Patent No. 6166290
; GENERAL INFORMATION:
; APPLICANT: Rea, Philip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-1202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-567-2991
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-972-927-1

alignment_scores:
Quality: 77.50 Length: 85
Ratio: 1.615 Gaps: 2
Percent Similarity: 56.471 Percent Identity: 28.235

alignment_block:
US-09-676-718A-1 x US-08-972-927-1/rev ..
Align seg 1/1 to reverse of: US-08-972-927-1 from: 1 to: 5232

```

```

32 GIupheSerSerGluAlaCysargGluLeuGlyPheSerSerAsnLeuLe 48
   ||||| ||||| : : ||||| : ||||| :
4414 GAATCTCTCGGATGCTCTGATGAGGCGCATCGTTCTAACATCTCA 4365
   ||||| ||||| : : ||||| : ||||| :
48 ucysSerSerCysAspLeuLeuGlyGlnPhe..... 58
   ||||| ||||| : : ||||| : ||||| :
4364 CAGCAGCAGCTCTTCATCAGAGCAGATCTTAGATCTCCGTAACAC 4315
   .....AsnLeuLeuGlnLeuAspProaspCysargGlyCys 70
   : : ||||| : : ||||| :
4314 GCACGTGAAGACTCAACAATTGCTCTCTCCACGCTGAATCTCTCC 4265
   71 CysGlnGlnGlnAlaGlnPheGlnThrLysLysLeuTyrAlaGlyAla 87
   || : : : : : : ||||| : : ||||| :
4264 TCCTCAGACAGCTCAGCATCAAGACCAAGAGATTTCTCGCATGTAT 4215
   87 eleuGluValCysGly**LysLeuGlyLysPheProGlnValGlnAla 104
   ||||| : ||||| ||||| ||||| : : : : :
4214 CCTTCAAGTGTG.....CCTCTCAGAGATTCACAAAGATCAGCATCA 4171
   104 heval 105
   : : |||
4170 TTGTG 4166

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-912-227-1

```

seq_documentation_block:
: Sequence 1, Application US/08912227
: Patent No. 5998171
: GENERAL INFORMATION:
: APPLICANT: Yu, Guo-Liang
: APPLICANT: Ni, Jian
: APPLICANT: Rosen, Craig A.
: TITLE OF INVENTION: Human Endokine Alpha
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C
: STREET: 1100 NEW YORK AVE., NW, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/912.227
: FILING DATE: Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,058
: FILING DATE: 16-AUG-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Steife, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1488, 0470001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2340
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1849 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 53...559
: US-08-912-227-1

```

```

alignment_scores:
  Quality: 75.00      Length: 98
  Ratio: 1.562
  Percent Similarity: 48.980  Percent Identity: 32.653

```

alignment_block:
US-09-676-718a-1 x US-08-912-227-1 ..

Align seg 1/1 to: US-08-912-227-1 from: 1 to: 1849

```

5 ProSerGlyCysLeuValProAlaPheGlyLysArgLeuLeuAla.. 20
   ||| ||| ||| ||||| : : ||||| :
1536 CCCCTCATTTGCACCTGCTCAGCCTTAGAATGTATTCTCTATACGTC 1585
   21 .....ThrValLeuGlnAlaValSerA 28
   ||||| ||||| ||||| ||||| :
1586 TTATTAAGACTTAAACTTAGCAATTCACCTTTTCAGGAAGCATATTCC 1635
   28 laPheGlyAlaGluPheSerSerGluAlaCysargGluLeuGlyPheSer 44
   : : ||||| : : ||||| :
1636 CTTTACCCAGGTGAGCAGAGTAGAGCTACACAGATCTTCTCTTACC 1685
   45 SerAsnLeuLeuCysSerSerCysAspLeuLeuGlyGlnPheAsnLeu 61
   ||||| : ||||| ||||| ||||| :
1686 AGCACACTTTTCTTTTCTCTC.....CTGAA 1714
   61 uGlnLeuAspProaspCysargGlyCysGlnGlnGlnAlaGlnPhe 77
   ||| ||||| ||||| ||||| ||||| :
1715 TCAGGAGATCCA.....GGATGCTGTTCAGGCGCTTATCCACCA 1755
   78 GluThrLysLysLeuTyrAlaGlyAlaIleLeuGluValCys 91
   : : ||||| : ||||| ||||| ||||| :
1756 AATTCCCTCTTACCTTTCAGAGGCCCATCTTAGTCAAAAGT 1797
   seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-972-927-4

seq_documentation_block:
: Sequence 4, Application US/08972927
: Patent No. 6166290
: GENERAL INFORMATION:
: APPLICANT: Rea, Philip A
: APPLICANT: Lu, Yu-Ping
: APPLICANT: Li, Ze-Sheng
: TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
: TITLE OF INVENTION: PLANTS
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
: STREET: One Commerce Square, 2005 Market Street, 22nd
: STREET: Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: US
: ZIP: 19103-7086
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/972.927
: FILING DATE: 18-NOV-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/031,040
: FILING DATE: 18-NOV-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/061,328
: FILING DATE: 08-OCT-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Doyle Deary Ph.D., Kathryn

```

```

;
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-1202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5175 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-972-927-4

alignment_scores:
    Quality: 74.50      Length: 85
    Ratio: 1.585        Gaps: 2
    Percent Similarity: 55.294    Percent Identity: 25.882

alignment_block:
US-09-676-718A-1 x US-08-972-927-4/rev ..

Align seg 1/1 to reverse of: US-08-972-927-4 from: 1 to: 5175

32 GluPheSerSerGluAlaGlnPheGluLeuGlnPheSerSerAsnLeuLe 48
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4385 GAATTCCTCTCGATGGCTTTGGATGAGAACATCACTTCTTACGTCA 4336
48 uCySerSerCysAspLeuLeuGlnPheAsnLeuLeuAsp 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4335 CTGCAGCAGTTGCTTCACAGAACAGTATCTAGATCTCGTAACAT 4286
65 roAspCysArgGly.....Cys 70
|||||
4285 GCAGCTGCACACATCAACACTGCTCTGCCACACGCAATCTCTCC 4236
71 CysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyralaGlyAla 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4235 TGCCCTGATTACCTCAGCATCAGCAGCAAGATTCTCGCGATAGTAT 4186
87 LeuGlnValCysGly**LysLeuGlyArgPheProGlnValGlnAla 104
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4185 CTTCAGATGTCCTCCCTCTCAG.....AGATCCACAGAGATCGCGCTG 4142
104 heVal 105
|||
4141 TTGTG 4137

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-441-139-6

seq_documentation_block:
; Sequence 6, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wiltrop, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCUDLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 441..2429
; US-08-441-139-6

alignment_scores:
    Quality: 74.00      Length: 94
    Ratio: 1.396        Gaps: 3
    Percent Similarity: 56.383    Percent Identity: 27.660

alignment_block:
US-09-676-718A-1 x US-08-441-139-6 ..

Align seg 1/1 to: US-08-441-139-6 from: 1 to: 2574

53 AspleuGlnGlnPheAsnLeuLeuGlnLeuAspProAspCysArg 69
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1881 AACCTTCTGGAAATTTGACCTTGCTGATTCCTCTGCCCTCGTGG 1930
69 yCyScsGlnGlnAlaGlnPheGluThrLysLysLeuTyralaGly 86
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1931 TGTTCGCCCAATTGAAATGACGTTGAAGTCGAT.....GCCAATG 1971
86 LaileuGlnValCysGly**LysLeuGlyArgPheProGlnVal 101
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1972 GTGTTTACTGTTTCAGCCGTCACAGAGCTCGTAAGGTAAGCTGAG 2021
102 GlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 118
||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2022 AAGCTTGTATCAGATGACAAAGGTGTTG..... 2054
118 eLysTyValArgGlySerAspProValLeuLysLeuLeuAspAsp 135
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2055 .....TCTGAGAAAGATATGAGCGCATGCTTAAGAGCGCG 2091
135 LyAsnLeuAlaGlnLeuSerLeuLys 145
|:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2092 AAGATTCGCTGAGAGATAGATTGTAAG 2123

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-178-477B-42

seq_documentation_block:
; Sequence 42, Application US/08178477B
; Patent No. 5756343
; GENERAL INFORMATION:
; APPLICANT: WU, CARL, CLOS, JOACHIM.
; APPLICANT: WESTWOOD, J. TIMOTHY.; RABINDRAN, SRIDHAR
; TITLE OF INVENTION: CELL STRESS
; TITLE OF INVENTION: TRANSCRIPTIONAL FACTORS
; NUMBER OF SEQUENCES: 42

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/178-477B
FILING DATE: 07-JAN-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/617,910
FILING DATE: 26-NOV-1990
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPPI
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4103US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-4800
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2781
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-178-477B-42

alignment_scores:
Quality: 69.50      Length: 167
Ratio: 0.827        Gaps: 9
Percent Similarity: 50.299   Percent Identity: 27.545

alignment_block:
US-09-676-718a-1 x US-08-178-477B-42/rev ..

Align seg 1/1 to reverse of: US-08-178-477B-42 from: 1 to: 2781

13 PhgGLyLysArgLeuLeuAlaThrValLeuGlnAlaValSer.... 27
|||||
1919 TTCTCAGCCGCACTGTATTGCGCACAGCCATGCCCTCTCCACAGATGC 1870
28 ....AlaphGlyAlaGluPheSerSerGluAlaGlyArgGluLeuGlyP 43
|||||
1869 TTCCGCGCTTGCGCGCTGCTGCATGTCGAGGCC.....TGATTGGCAT 1826
43 heSerSerAsnLeuLeuGlySerSerCysAspLeuLeuGlyGlnPheAsn 59
|||||
1825 CCAGCAGCGCTGTTGGTGGCGACATCGAAGTGTGCGCA.....AAC 1782
60 LeuLeuGlnLeuAspProAspCysAlaGly..CysCysGlnGlu.. 74
|||||
1781 TTGCCCGCTGCTCACTGTCGATGCCCAAAATTTTGGTGGTGGTCCAGTTT 1732
75 ....AlaGlnPheGluThrLysLysLeuTyrAlaGlnAlaIleLeuGlu 90
|||||
1731 GGTGCGCTTTCACAGCCGCTCGGGAAGCAGAGAGCTTCACAAAATATCCG 1682
90 aLcysGly**LysLeuGlyArgPheProGlnValGlnAlaPhe..... 104
|||
1681 TCGGCATATTCCTCCTCATGATGAAGTTCGACACTGTACAGATGTTCCGGCGCA 1632
105 ...ValArgSerAspLysProLysLeuPheArgGly..... 115

```

```

1631 GTCATTAGATCATCATCGGAGACTGTTCTCCGCGTCAGACTTATAT 1582
116 .....Leu GlnIleLysTyrValArgGlySerAspPro 126
1581 CATTTGGCTGAGCAATCATCATGAGTTGCTGCTGCTGATGC..... 1538
127 ValLeuLysLeuLeuAspAspAsnGlyAsnIleAlaGluLeuSerIle 143
1537 .....TTGTCGCGT 1530
143 eleuLysTrpAsnThrAspSerValGluGluPheLeuSerGluLysLeu 159
1529 TGGCGCTTGGCGCAGCAATCATCATGAGATATCTCCACATCAAGTA 1481

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-441-139-1

seq_documentation_block:
Sequence 1, Application US/08441139
Patent No. 5773245
GENERAL INFORMATION:
APPLICANT: Wiltrop, Dr. Karl D.
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,139
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,997
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8646
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2780 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 285..2333
US-08-441-139-1

alignment_scores:
Quality: 69.00      Length: 143
Ratio: 1.015        Gaps: 5
Percent Similarity: 47.552   Percent Identity: 23.776

alignment_block:
US-09-676-718a-1 x US-08-441-139-1 ..

```

Align seg 1/1 to: US-08-441-139-1 from: 1 to: 2780

```

53 AspleuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCysArgI 69
   ::::::::::::::::::::|||
1773 AATCATATAGAGTAAGTTGATTAACCGGCAATCCACGACCAAGAG 1822
   ::::::::::::::::::::|||
69 yCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 86
   |||
1823 TGTACCTCAAAATTTGAAGTGCACATTT.....GCACCTTGACGCTAATG 1863
   |||
86 LalleleuGlnValCysGly**LysLeuGlyArgPheProGlnValGln 102
   ::::::::::::::::::::|||
1864 GATATCTGAAGGTGTCTGCCACACATTAAGGACGTAATCCGAAATCT 1913
   ::::::::::::::::::::|||
103 AlapheValArgSerAspLysProLysLeuPheArgGlyLeuGlnIle 119
   ::::::::::::::::::::|||
1914 ATCACCATACATACGATTAAGGTAGATTAAACCAAGAAAGATGATG 1963
   ::::::::::::::::::::|||
119 sTyArgValArg.....GlySerAspProValLeuLysL 130
   |||
1964 AATGGTTGAAGAGCGCTGAATAATTCGCTTGAGACGCTCTATCAAG 2013
   ::::::::::::::::::::|||
130 eu.....LeuAspAsp..... 133
   |||
2014 CCAAGGTGAATCTGAAACAAATTAGAAACTACGCTCCTCTTGAAA 2063
   |||
134 .....AsnGlyAsnIleAlaGlnGlnLeu..... 141
   |||
2064 AACCAAGTTAATGTGACCTAGTGAAATAATGGAAGAAAGACAGAGA 2113
   |||
142 .....SerIleLeuLysTrpAsnTrpAspSerV 151
   |||
2114 AACCTTATATAGATGCTGTACACGATGTTTATGATGTTAGATTAAC 2163
   |||
151 alGlnGlnPheLeuSerGlnLysLeuGln 160
   |||
2164 TTGAACCGCCATGTGCTGAGACCTTGAT 2192

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-103-840A-1

```

seq_documentation_block:
: Sequence 1, Application US/09103840A
: Patient No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103, 840A
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

alignment_scores:

```

Quality: 68.00 Length: 141
Ratio: 0.883 Gaps: 6
Percent Similarity: 54.610 Percent Identity: 26.241

```

alignment_block:

US-09-676-718A-1 x US-09-103-840A-1 ..

Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529

```

12 AlapheGlyLysArgLeuLeu.....LeuAl 20
   ::::::::::::::::::::|||
502397 ACCTTCGCCCTCCGCGCTCATATGAGTACTGGGGGTGCCAACCAATCTGGC 502446
   ::::::::::::::::::::|||
20 aThrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheSerGln 37
   ::::::::::::::::::::|||
502447 GGTGTAGAGCGAGGCTCTGATCGCTCAAGTCCAGAGTGCAGACCGTGC 502496
   ::::::::::::::::::::|||
37 LacArgGlnLeuGlnPhe.....SerSerAsnLeuLeuCysSer 51
   |||
502497 CCATACGCCGGGTTCACGCCGACGGGGGAACCGGCTGTCTC..... 502537
   ::::::::::::::::::::|||
52 CysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCysAr 68
   |||
502538 ...GACCTGCTCAACCGCGCTCGCATCACACCGCTA...CCCAACACCGC 502581
   ::::::::::::::::::::|||
68 gGlyCysCysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 85
   |||
502582 GGGGTCCCGCACCGCGCGGGAAGCGGTCTCTGACACCGCAGTGGCCCGTG 502631
   |||
85 LAlaIle.....LeuGlnValCysGly**LysLeu 95
   |||
502632 AGCGCCTGAACACCAACTGCTGCAAGCTGAGGTATTCGACGACGACCC 502681
   |||
96 GlyArgPheProGlnValGlnAlaPheValArgSerAspLysProLys 112
   ::::::::::::::::::::|||
502682 ACCCTGTGCGCTGATGCGGCTCGAATTAGTCCGCGCTGCAGAACAAATGGT 502731
   |||
112 uPheArgGlyLeuGlnIleLysTyArgValArgLysSerAspProVal 128
   ::::::::::::::::::::|||
502732 GAGACGAGGATTTGTGTCTTACCTGACACCAACGACGACCGGTGTGG 502781
   |||
129 ..LysLeuLeuAspAspAsnGly 135
   ::::::::::::::::::::|||
502782 CCCGCCGCTAGAAATACCGCT 502804

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-752-760A-1

```

seq_documentation_block:
: Sequence 1, Application US/08752760A
: Patient No. 5877011
: GENERAL INFORMATION:
: APPLICANT: Armentano, Donna
: APPLICANT: Gregory, Richard J.
: APPLICANT: Smith, Alan E.
: TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Baker & Bots, L.L.P.
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: NY
: COUNTRY: U.S.A.
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/752,760A
: FILING DATE: 20-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Seide, Rochelle K
: REGISTRATION NUMBER: 32,300
: REFERENCE/DOCKET NUMBER: A31385
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-705-5000

```


TELEFAX: 212-705-5020
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 35081 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-752-760A-1

alignment_scores:
 Quality: 67.00 Length: 94
 Ratio: 1.241 Gaps: 5
 Percent Similarity: 57.447 Percent Identity: 31.915

alignment_block:
 US-09-676-718A-1 x US-08-752-760A-1/rev ..

Align seg 1/1 to reverse of: US-08-752-760A-1 from: 1 to: 35081

```

30 GYAlaGluPheserSerglAlaCysArgLulLeuGlyPheserSerg 46
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3750 GCGCAGCAGTCTCCCTCG..GGCAGTGAAGAACTGGCAGAGGGGCT 3704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 nLeuLeu.CysSerSergCysAspLeuLeuGlyLnpheanLeu..... 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3703 ACTGCTGCTGCTGGCAGTACCTCGAGCCAGCTGCTGCTGCTATG 3654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 ..LeuGluLeuAspProAspCysArgGlyCysGlnLulLulAla 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3653 CGGCTCGGCTG...CGCGGCTCGCGGCTGCTCAACGAGAGTTC 3607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
76 nPheGluThrLysLysLeuTyraLalLeuGluValCysGly* 93
   : ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3606 CCACAGGTGCGCATGTCAGTGCAGGAATTGCTGAGACACTGGGC. 3558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
93 **LysLeuGlyArgPheProGlnValGlnAlaPheValArgSerAsp 109
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3557 .....GCCCTCCACCGTCATCCATCCATTCCTGACGAACTCGG 3516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110 ProlLysLeuPheArgGlyLeuGlnLleLys 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3515 CCCATCTGGCAGCGGCTGTCAATTAAG 3486

```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-605-150A-3

seq_documentation_block:

Sequence 3, Application US/08605150A
 Patent No. 6103520
 GENERAL INFORMATION:
 APPLICANT: Topfer, Reinhard
 APPLICANT: Hausmann, Ludwig
 APPLICANT: Schell, Jozef
 TITLE OF INVENTION: GLYCEROL-3-PHOSPHATE DEHYDROGENASE
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klein & Szekeres
 STREET: 4199 Campus Drive, Suite 700
 CITY: Irvine
 STATE: CA
 COUNTRY: USA
 ZIP: 92715
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/605.150A
 FILING DATE: 01-MAR-1996
 CLASSIFICATION: 800

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/EP94/02936
 FILING DATE: 02-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DE P4329827.3
 FILING DATE: 03-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Szekeres, Gabor L.
 REGISTRATION NUMBER: 28, 675
 REFERENCE/DOCKET NUMBER: 542-04-PA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 714-854-5502
 TELEFAX: 714-854-4897
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1464 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Cuphea lanceolata
 IMMEDIATE SOURCE:
 LIBRARY: ZAP cDNA library
 CLONE: C1GPDH109
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 45..1187
 US-08-605-150A-3

alignment_scores:
 Quality: 66.50 Length: 87
 Ratio: 1.255 Gaps: 5
 Percent Similarity: 60.920 Percent Identity: 33.333

alignment_block:
 US-09-676-718A-1 x US-08-605-150A-3/rev ..

Align seg 1/1 to reverse of: US-08-605-150A-3 from: 1 to: 1464

```

20 AlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheserSerg 36
   ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 GGCATCTTTGCAAAAGCCCTCAGCAGACTTTCTGTTCTTCCGCCCAAC 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 uAlaCysArgGluLeuGlyPheserSergAsnLeuLeuCysSerSerg 52
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 AGTTGTGATGATGATCAGCAGCTCCGACGCTTCGAAAAAAGTAGTGTCC 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53 .....AspLeuLeuGlyGln...PheAsnLeuLeuGlnLeu 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 TTAACAGATGAAACAAACCTTGAGAGACGTTCTATCTCCGACGCC 825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 AspPro...AspCysArgGlyCysGlnGluGluAlaGluPheGluThr 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 GATCTCATTAATGCGCCTTTGTGTTGTTCCATCTCCAAATCCATCA 775
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
79 rLys...LysLeuTyraLalLeuGluValCysGly**LysL 95
   :||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
774 CAAAACCCGCTGATAGGCCCAATATTTCTTCA...TGTCCACAAAGC 728
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 enuLysArgPhe 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
727 TCTACTCTCTTC 717

```

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-942-686-1

seq_documentation_block:

Sequence 1, Application US/08942686
 Patent No. 6183988
 GENERAL INFORMATION:

```

? APPLICANT: Bloch, Donald B.
? TITLE OF INVENTION: LEUKOCYTE-SPECIFIC PROTEIN AND GENE, AND
? TITLE OF INVENTION: METHODS OF USE THEREOF
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNER, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
? STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
? CITY: WASHINGTON
? STATE: D.C.
? COUNTRY: USA
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/942,686
? FILING DATE: 02-OCT-1997
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/027,347
? FILING DATE: 02-OCT-1996
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: GOLDSTEIN, JORGE A.
? REGISTRATION NUMBER: 29,021
? REFERENCE/DOCKET NUMBER: 0609.4330001/JAG/BJD
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-371-2600
? TELEFAX: 202-371-2540
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2905 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 107..2365
? US-08-942-686-1

alignment_scores:
  Quality: 66.50      Length: 133
  Ratio: 1.023        Gaps: 7
  Percent Similarity: 48.872   Percent Identity: 24.060

alignment_block:
US-09-676-718A-1 x US-08-942-686-1 ..
Align seg 1/1 to: US-08-942-686-1 from: 1 to: 2905
34 SerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1985 AGCCAAAGTGTGTGAGGAATCTGAGTCTCTGAGAGGAGAGATGTGTC 2034
50 rSerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAsp 67
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2035 TGAGGAACAGTTGAATGTGAGTCTCTCTTGAAGTC..... 2074
67 ySaArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyr 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2075 ....TATGTGCTG...TCTGAGAGCTCTTTTGGCAAGATTCATTCAC 2116
84 AlaGlyAlaIleLeuGluValCys..Gly**Lys.....Le 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2117 TATATTTATTATAGAGAGCGGTCTCAAGGCTGAAGGAGGCCATGTGTT 2166
95 uGlyArg.....PheProGlnValGlnA 103
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2167 GGATAAAATCAAGAAAGGCTGATGACACGCGTTACCCCAAGTGGAG 2216
103 lAphValArgSerAspLysProLysLeuPheArgGly..... 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2217 GGTGTGTACAA...GACATGGCGCTCATCTTCACAAACACAGGCGCTCT 2263
116 .....LeuGlnIleLysTyr 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2264 TACAGATACAGAGATTTGGCCAAATGGGATTTAGACTGGAGGCGTGT 2313
120 rValArgGlySerAspProValLeuLysLeuLeuAspAspAsnGlyAsn 136
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2314 TGAGGAAGATTTCAAGGAAGTGTTCATTTCAGGAACAAATGGGAAC 2362
seq_name: /cgn2.6/prodata/2/ina/6B_COMB.seq:US-09-180-439-1
seq_documentation_block:
? Sequence 1, Application US/09180439
? Patent No. 625532
? GENERAL INFORMATION:
? APPLICANT: Dixon, Mark S
? APPLICANT: Hatzixanthis, Kostas
? APPLICANT: Jones, David A
? APPLICANT: Jones, Jonathan DG
? TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
? FILE REFERENCE: 620 - 53
? CURRENT APPLICATION NUMBER: US/09/180,439
? EARLIER FILING DATE: 1998-12-06
? EARLIER APPLICATION NUMBER: PCT/GB97/01249
? EARLIER FILING DATE: 1997-05-08
? EARLIER APPLICATION NUMBER: GB 9609681.3
? EARLIER FILING DATE: 1996-05-09
? EARLIER APPLICATION NUMBER: GB 9619924.5
? EARLIER FILING DATE: 1996-09-24
? NUMBER OF SEQ ID NOS: 10
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 1
? LENGTH: 3979
? TYPE: DNA
? ORGANISM: Lycopersicon esculentum
? US-09-180-439-1

alignment_scores:
  Quality: 66.00      Length: 179
  Ratio: 0.857        Gaps: 7
  Percent Similarity: 43.017   Percent Identity: 21.788

alignment_block:
US-09-676-718A-1 x US-09-180-439-1/rev ..
Align seg 1/1 to reverse of: US-09-180-439-1 from: 1 to: 3979
14 GlyLysArgLeuLeuLeuAlaThrValLeu.....GlnAlaValSe 27
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2854 GGTTCCTCATTTGTTTATTCACACTGTCTCTCATCCCTTCAAAATGTTCAA 2805
27 rAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuGlyPhe. 43
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2804 TTAGACTCGTGTGTAGTCTTTCGAGAAATGATTCAGCCCTGATGATCTTATGA 2755
44 .....SerSerAsnLeuLeuCysSerSerCysAspLeuGlyGln 57
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2754 TTCGAGATACAGAAACATGATTCAGCCCTGATGATCTTATGTCGA 2705
58 PheAsnLeuLeuGln..... 62
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2704 TGCATTTTATTCGATGTCACTTAAACTCTCAGCTGTGCAAAAGTTC 2655
63 .....LeuAspProAspCysArgG 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2654 CAACCAATCGGAATGTGTGCTGATGATGATGTTGCTCTCTTAATCAAGA 2605

```

```
69  lYcyscysglngluaglaphgluhrlyslsleuTyralagly 85
    |||
    : : : : :
2604  CTTCG.....ACGTTTTCGCAATGTGCCAAGACGAGGATTCATCT 2561
    |||
    : : : : :
86  AlaIleuGlulValCysgly**LysLeu..... 95
    |||
    : : : : :
2560  GCTAGTTCATGTCGCAAGTTCAGACTTATTCAGTGAACATCCATGCT 2511
    |||
    : : : : :
96  .....GlyArgPheProGluValGlnAlaPheValArgSer.... 107
    |||||
    : : : : :
2510  AAAATTGTTGGAAAGAGTCCAGAAAGTTGTTATTCGATATCAAAAA 2461
    |||
    : : : : :
107  ..... 107
2460  CCTGAGGCTACTAATATTCGCAAAAAATGTGTATTCCTCCCTCCAGA 2411
108  .....AsplyProLysLeuPheArgGlyLeuGlnIleLysTyra 121
    |||
    : : : : :
2410  TTGTTCTTCGCCAAATCAAGTATTTTATGATGATTAATGGAAATGAGA 2361
121  lArgGlySerAspProValLeuLysLeuLeuAspAsp 133
    : : : : :
2360  TGAAGGAGCTCTCCT...CTGAACATATTAATGATGAC 2327

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-180-439-2

seq_documentation_block:
; Sequence 2, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; EARLIER FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-2

alignment_scores:
    Quality: 66.00      Length: 179
    Ratio: 0.857        Gaps: 7
    Percent Similarity: 43.017    Percent Identity: 21.788

alignment_block:
US-09-676-718a-1 x US-09-180-439-2/rev ..
Align seg 1/1 to reverse of: US-09-180-439-2 from: 1 to: 3979

14  gLylySArgLeuLeuAlaThrValLeu.....GlnAlaValSe 27
    |||||
    : : : : :
2854  GGTTCCCTCATGTTTATCAACTGTCCTATCCCTTTCAAATGTCAAA 2805
27  rAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuGlyPhe. 43
    : : : : :
2804  TAGACTCGTGTAACTGTGCGAATGATCGAGAGAGATCTATGA 2755
44  .....SerSerAsnLeuLeuCysSerSerCysAspLeuGlyGln 57
    |||||
    : : : : :

```

```
2754  TTGCGAGATCAGGAACATGATTTACGCCCTCATGATCTTAAAGTCCA 2705
58  PheAsnLeuLeuGln..... 62
    |||||
    : : : : :
2704  TGCATTTATTCATGTCACCTTAAACCTCAGCTCTGGCAAGTCC 2655
63  .....LeuAspProAspCysArg 69
    |||||
    : : : : :
2654  CAACCATGCGAAATGCTGCTTGAGTTGATGATGTCCTCAAAATCAAGAA 2605
69  lYcyscysglngluaglaphgluhrlyslsleuTyralagly 85
    |||
    : : : : :
2604  CTTCG.....ACGTTTTCGCAATGTGCCAAGACGAGGATTCATCT 2561
    |||
    : : : : :
86  AlaIleuGlulValCysgly**LysLeu..... 95
    |||
    : : : : :
2560  GCTAGTTCATGTCGCAAGTTCAGACTTATTCAGTGAACATCCATGCT 2511
96  .....GlyArgPheProGluValGlnAlaPheValArgSer.... 107
    |||||
    : : : : :
2510  AAAATTGTTGGAAAGAGTCCAGAAAGTTGTTATTCGATATCAAAAA 2461
107  ..... 107
2460  CCTGAGGCTACTAATATTCGCAAAAAATGTGTATTCCTCCCTCCAGA 2411
108  .....AsplyProLysLeuPheArgGlyLeuGlnIleLysTyra 121
    |||
    : : : : :
2410  TTGTTCTTCGCCAAATCAAGTATTTTATGATGATTAATGGAAATGAGA 2361
121  lArgGlySerAspProValLeuLysLeuLeuAspAsp 133
    : : : : :
2360  TGAAGGAGCTCTCCT...CTGAACATATTAATGATGAC 2327

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-180-439-7

seq_documentation_block:
; Sequence 7, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatzixanthis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; EARLIER FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4123
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-7

alignment_scores:
    Quality: 66.00      Length: 179
    Ratio: 0.857        Gaps: 7
    Percent Similarity: 43.017    Percent Identity: 21.788

alignment_block:
US-09-676-718a-1 x US-09-180-439-7/rev ..
Align seg 1/1 to reverse of: US-09-180-439-7 from: 1 to: 4123

```

14 Glytysargleuleuleumalaphvalleu.....GlnalaValse 27
|||:::|||||:::|||||:::
2998 GGTTCCTCCATTGGTTATTTCACATCTCTCCATCCCTTCAATATTCAAA 2949
27 rAlapheglialagiupheserSerglualcysargluleulellyphe. 43
: :::: ||| ||||| :::::
2948 TAGACTCGTTGGTAAGTCTTGGCAGCAATGCATTTGCAGAGAGATCTATGA 2899
44SerSerSnuleuleucysSerSercysAspIeuleuGlyGln 57
|||:::|||||:::|||||:::
2898 TTCGAGATACAGAAACAAATGATTTCCGCCCTGATGATCTATAGGTCA 2849
58 pheAsnIeuleuGln..... 62
|||||:::|||||:::
2848 TGCAAATTATTTCGATGTCAACCTTAAACTCTCAAGCTCGCAAAAGTCC 2799
63LeuAspProAspCysArg 69
|||:::|||||:::
2798 CAACCCATAGGAAATGTCCTGGTTCAGTTGATCTTCCTCAATACCAAGA 2749
69 lYcscGsgInglIngluInlagInphegluInhrlyslsyleuTyAlaGly 85
||| :::: |||:::|||||:::
2748 CTTCG.....AGCTTTTGGCAATGTCCAAAGACCGAGGATTTCACT 2705
86 AlaIleleuGluValCysGly**LysLeu..... 95
||| ||| |||:::
2704 GCTAGTTCATTGCATGCAGAAAGTTGAGACTTTCAGTGAACATCCAAATGCT 2655
96GlyArgpheProGlnValGlnAlaPheValArgSer... 107
||| ||||| |||::: ||| |||
2654 AAAATTTCCTGGAAAGAGCCCGCAAGAAAGTTGTATTCTGCATATCAAAA 2605
107 107
2604 CCTGGAGGCTACTAATATTGCCAAAAAATTGTGGATTGCTCCCTCCAGA 2555
108AspLysProLysIleuPheArgIleuGlnIleuTyAla 121
||| :::: ||| :::::
2554 TTGTTTCTGCCAAAAATCAAGATTTCCTTGGTATGATCTTAATTGAAATAGA 2505
121 lArgGlySerAspProValIleuLysLeuLeuAsp 133
|||||:::|||||:::
3504 TGAAGGAGCTCTCTCT...CTGAAACTATTTAGATGAC 2471

CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (hemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43926 to AAA43931 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX
 XX

OS Homo sapiens.
 XX
 XX Key
 XX CDS

Location/Qualifiers
 5..493
 /*tag- a
 /note- "the coding region is also claimed in
 in claim 13"

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718A-1 x AAA44484 ..

Align seg 1/1 to: AAA44484 from: 1 to: 626

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
 28 AVGGGGGTGGGCCGAGTGGGTGTGGTGGCGGGTGGGCTACGGTT 77
 17 UleuDeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
 78 GTGTGGCGACTGTGCTTCAAGCGGTGTCTCTTTGGGGCAGAGTTT 127
 34 erSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGlySer 50
 128 CATCGAGAGCATGACAGAGATTAGCTTTTCAGCACTGCTTGCAGC 177
 51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCys 67
 178 TCTTGATCTCTCTCGACAGTCACTGCTTCACTGAGTCCGATGG 227
 67 sArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyrA 84
 228 CAGAGAGATGCTGTGAGAGAGACCAATTGAAACCAAAAGCGTATG 277
 84 IaGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
 278 CAGGAGCATCTCTGAACTTGTGATGAAATGGGAAGTCCCTCAA 327
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
 328 GTCCAGCTTTGTAGAGATGATTAACCAACTGTTCAAGAGACTGCA 377
 117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
 378 AATCAAGATATGCGGTGTGACAGACCTGTATTAAGCTTTTGGACGACA 427
 134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrPAsnThrAspSer 150
 428 ATGGGAAATGCTGTAAGAACTGACCAATTCCTCAATGGAACACAGACAGT 477
 151 ValGluGluPheLeuSerGluLysLeuGluArgIle 162
 478 GTAGAGAAATCTCTGACGATAAAGTTGGAAACGCATTA 513

seq_name: /SIDS1/gcgdata/hold-1/geneseq/geneseq-emb1/NA2000.DAT:AA234464

seq_documentation_block:

ID AA234464 standard; cDNA; 1244 BP.

AC AA234464;

DT 01-FEB-2000 (first entry)

DE Human 15 kDa selenoprotein cDNA.

KW Selenoprotein; selenium; differential expression; tumour; human;
 KW prostate cancer; diagnosis; polymorphism; gene therapy; ss.
 XX

W09951637-A1.
 14-OCN-1999.
 06-APR-1999: 99WO-US07560.
 06-APR-1998: 98US-0080850.
 (USSR) US DEPT HEALTH & HUMAN SERVICES.
 Gladyshev VN, Woolton JC, Hatfield DL, Jeang K;
 WPI: 2000-013045/01.
 P-PSDB; AAY32112.

New isolated selenoprotein polypeptides, used to develop products for
 detecting susceptibility to or treating cancers e.g. prostate cancers

Claim 13: Page 55-56; 67pp; English.

The present sequence represents a human cDNA sequence encoding a
 15 kDa selenoprotein (see AAY32112). The coding region (also
 claimed) includes an in-frame TGA codon encoding selenocysteine. A
 conserved stem-loop structure (see also AA234473) in the 3'
 untranslated region (3'UTR) of the corresponding mRNA is essential
 for recognition of this UGA as a codon for selenocysteine. This
 structure is termed a stem-loop selenocysteine insertion sequence
 (SECIS) element. 2 Polymorphisms have been detected in the 3'UTR
 of the human selenoprotein cDNA, one of them located within the
 SECIS. There is a correlation between the presence of these
 polymorphisms and the presence of cancer, including prostate cancer.
 The polymorphisms are more prevalent in the African American
 population. The determination of an individual's genotype may be
 used as an indicator of the need for dietary selenium
 supplementation to inhibit tumour development. Probes and primers
 useful for detecting polymorphisms in the cDNA are disclosed. Also
 provided are compositions and methods for determining the effect of
 candidate tumour therapeutics on the expression of the 15 kDa
 selenoprotein. Gene therapy can be used to treat tumors by
 overexpressing the 15 kDa selenoprotein in cells having an
 abnormally low amount of 15 kDa selenoprotein, or in the cells of
 high-risk patients. The 15 kDa selenoprotein gene maps to human
 chromosome 1, approximately to 1p31.

Sequence 1244 BP; 380 A; 217 C; 257 G; 390 T; 0 other;

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x AAX34464 ..

Align seg 1/1 to: AAX34464 from: 1 to: 1244

```

1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||||
5 ATGGCGGCGTGGCCGAGTGGTGTCTGGTCCGCGCTTTGGCGTACGCTT 54
  |||||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlyPheS 34
  |||||
55 GTTGTGGGAGCTGTGCTTCAAGCGGTGTCTGTTTGGGCGAGACTTTT 104
  |||||
34 eSerGlyAlaCysArgGlyLeuGlyPheSerSerAsnLeuLeuCysSer 50
  |||||
105 CATCGAGGACATGCAGAGAGTTAGGCTTTTCTAGCAACTGTGCTTGCAC 154
  |||||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuGlnLeuAspProAspCys 67
  |||||
155 TCTTGTGATCTTCTCGACAGTTCACACTGCTTCACTGATCTGATG 204
  |||||
67 sArgGlyCysCysGlnGlyLeuAlaGlnPheGlyThrLysLysLeuTyr 84
  |||||
205 CAGAGGATCTGTCAAGAGGAAAGCAACATTTGAAACCAAAAGCTGTATG 254
  |||||
84 IAGLIAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
  |||||
255 CAGGAGCATTTCTTGAAGTTTGTGATGAATTTGGGAAGTTCCTCA 304
  |||||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
  |||||
305 GTCCAGACTTTGTAGAGAGTGAATACCAACCTTTCAGAGGACTGCA 354
  |||||
117 nIleLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAs 134
  |||||
355 AATCAAGATGTGCGGTTCAGACCTGTATTAAGCTTTGGAGGACA 404
  |||||
134 sNGlyAsnIleAlaGlyLeuSerIleLeuLysTrpAsnThrAspSer 150
  |||||
405 ATGGAAACTTGTCTGAAGAACTGAGCATTCTCAAAATGGAACAGACAG 454
  |||||
151 ValGlnGluPheLeuSerGlyLysLeuGluArgIle 162
  |||||
455 CTGAAAGAAATTCCTGAGTCAAAAAGTTGGAACCATTA 490
  |||||

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NAL199. DAT: AAX00680

seq_documentation_block:

ID AAX00680 standard; DNA; 1563 BP.

AC AAX00680;

DT 25-MAR-1999 (first entry)

XX Human secreted protein gene 70 clone HMSGT42.

XX Human: secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; fetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX OS W09842738-Al.

XX 01-OCT-1998.
 PD 19-MAR-1998; 98WO-US05311.
 PF
 XX

PR 30-MAY-1997; 97US-0050937.
 PR 21-MAR-1997; 97US-0041276.
 PR 21-MAR-1997; 97US-0041277.
 PR 21-MAR-1997; 97US-0041281.
 PR 21-MAR-1997; 97US-0042344.
 PR 30-MAY-1997; 97US-0048069.
 PR 30-MAY-1997; 97US-0048094.
 PR 30-MAY-1997; 97US-0048095.
 PR 30-MAY-1997; 97US-0048096.
 PR 30-MAY-1997; 97US-0048099.
 PR 30-MAY-1997; 97US-0048131.
 PR 30-MAY-1997; 97US-0048135.
 PR 30-MAY-1997; 97US-0048154.
 PR 30-MAY-1997; 97US-0048160.
 PR 30-MAY-1997; 97US-0048186.
 PR 30-MAY-1997; 97US-0048187.
 PR 30-MAY-1997; 97US-0048188.
 PR 30-MAY-1997; 97US-0048350.
 PR 30-MAY-1997; 97US-0048351.
 PR 30-MAY-1997; 97US-0048352.
 PR 30-MAY-1997; 97US-0048353.
 PR 05-AUG-1997; 97US-0054804.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Duan R, Edner R, Ferrle AM, Florence KA;
 PI Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS;
 PI Rosen CA, Ruben SM, Shi Y, Young P;

DR MPI: 1999-070066/06.
 DR P-PSDB: AAM67876.

XX New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX

PS Claim 1; Page 230-231; 385pp; English.

XX This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAX00602) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 87 novel genes and their fragments (nucleic
 CC acid sequences: AAX00611-X00724; amino acid sequences AAM67807-W68004)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 87
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX00611 for described uses).

XX Sequence 1563 BP; 478 A; 252 C; 301 G; 531 T; 1 other;

alignment_scores:

Quality: 819.00 Length: 162
 Ratio: 5.119 Gaps: 0
 Percent Similarity: 98.765 Percent Identity: 98.765

alignment_block:

US-09-676-718a-1 x AAX00680 ..

Align seg 1/1 to: AAX00680 from: 1 to: 1563

1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17

```
|||||
40 ATGGCGGCTGGGCGAGTGGGTGTGTCGGCGGCGTTGGGCTACGGTT 89
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheS 34
90 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGGTTT 139
34 eSerGluAlaCysArgGluLeuGlnPheSerSerAsnLeuLeuCysSer 50
140 CATCGAGGCAATGCGAGAGATTAGGCTTTTACCACTTCTTGGCAGC 189
51 SerCysAspLeuLeuGlnPheAsnLeuLeuGlnLeuAspProAspCys 67
190 TCTTGTGATCTTCTCGGACAGTTCACTGCTTCACTGATGATCTGATG 239
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGluThrLysLysLeuTYR 84
240 CAGAGGATGCTGTGAGGAGAACCAATTTGAAACCAAAAGCTGTATG 289
84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
290 CAGAGCTATTCTTGAAGTTGTGATGAAATTTGGGAGGTTCCCTCA 339
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
340 GTCCAGCTTTTGTAGAGATGAATTAACCACTGTTCAAGAGCACTCA 389
117 nIleLysTyValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
390 AATCAAGTATGTCCTGCTGAGACCCGTATTAAGCTTTTGGACGACA 439
134 sngLysAsnIleAlaGluGlnLeuSerIleLeuLysTrpAsnThrAspSer 150
440 ATGGGAACATGCTGAAGAACTGAGCAATTCCAATGGAACACAGACAGT 489
151 ValGluGlnPheLeuSerGluLysLeuGlnArgIle 162
490 GTAGAGAAATTCCTGAGGAAACTTGTGAACGCA 525
seq_name: /STDS1/gcgdata/mol-d-geneseq/geneseq-emb1/NA2000.DAT:AAF16335
seq_documentation_block:
ID AAF16335 standard; cDNA; 1637 BP.
AC AAF16335;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SMO ID NO:770.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnerrary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO20005174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
```

```
DR P-PSDB; AAB57132.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer
XX
PS Claim 1; Page 1189-1190; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytosolic,
CC cardioactive, immunomodulatory, muscular, vulnerrary, gastrointestinal,
CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX
SQ Sequence 1637 BP; 500 A; 273 C; 323 G; 541 T; 0 other;
```

alignment_scores:

Quality:	819.00	Length:	162
Ratio:	5.119	Gaps:	0
Percent Similarity:	98.765	Percent Identity:	98.765

alignment_block:

US-09-676-718A-1 x AAF16335 ..

Align seg 1/1 to: AAF16335 from: 1 to: 1637

```
1 MetaAlaAGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
99 ATGGCGGCTGGGCGAGTGGGTGTGTCGGCGGCGTTGGGCTACGGTT 148
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheS 34
149 GTTGTGGCGACTGTGCTTCAAGCGGTGTCTGCTTTGGGCGAGGTTT 198
34 eSerGluAlaCysArgGluLeuGlnPheSerSerAsnLeuLeuCysSer 50
199 CATCGAGGCAATGCGAGAGATTAGGCTTTTACCACTTCTTGGCAGC 248
51 SerCysAspLeuLeuGlnPheAsnLeuLeuGlnLeuAspProAspCys 67
249 TCTTGTGATCTTCTCGGACAGTTCACTGCTTCACTGATGATCTGATG 298
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGluThrLysLysLeuTYR 84
299 CAGAGGATGCTGTGAGGAGAACCAATTTGAAACCAAAAGCTGTATG 348
84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
349 CAGAGCTATTCTTGAAGTTGTGATGAAATTTGGGAGGTTCCCTCA 398
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
399 GTCCAGCTTTTGTAGAGATGAATTAACCACTGTTCAAGAGCACTCA 448
117 nIleLysTyValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
449 AATCAAGTATGTCCTGCTGAGACCCGTATTAAGCTTTTGGACGACA 498
134 sngLysAsnIleAlaGluGlnLeuSerIleLeuLysTrpAsnThrAspSer 150
499 ATGGGAACATGCTGAAGAACTGAGCAATTCCAATGGAACACAGACAGT 548
151 ValGluGlnPheLeuSerGluLysLeuGlnArgIle 162
```


|||||
549 GTAGAGAAATTCCTGAGTGAAGGTTGGAACGCTA 584

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ34468

seq_documentation_block:

ID AAZ34468 standard; cDNA; 1216 BP.

AAZ34468;

01-FEB-2000 (first entry)

Mouse 15 kDa selenoprotein cDNA.

Selenoprotein; selenium; differential expression; tumour; mouse;

prostate cancer; diagnosis; polymorphism; gene therapy; ss.

Mus musculus.

Location/Qualifiers

Key 11..490

CDS /tag= a

stem_loop 1049..1127

/tag= g /note= "stem-loop selenocysteine insertion sequence

(SECIS element)"

WO9951637-A1.

14-OCT-1999.

06-APR-1999; 99WO-US07560.

06-APR-1998; 98US-0080850.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Gladyshev VN, Wootton JC, Hatfield DL, Jeang K;

WPI; 2000-013045/01.

P-PSDB; AAY32113.

New isolated selenoprotein polypeptides, used to develop products for detecting susceptibility to or treating cancers e.g. prostate cancers

Claim 13; Page 58-59; 67pp; English.

The present sequence represents a murine cDNA sequence encoding a 15 kDa selenoprotein (see AAY32113). The coding region includes an in-frame TGA codon encoding selenocysteine. A conserved stem-loop structure (see also AAZ34474) in the 3' untranslated region (3'UTR) of the corresponding mRNA is essential for recognition of this UGA as a codon for selenocysteine. This structure is termed a stem-loop selenocysteine insertion sequence (SECIS) element. 2 CC polymorphisms have been detected in the 3'UTR of the human selenoprotein cDNA (see AAZ34464). There is a correlation between the presence of these polymorphisms and the presence of cancer. The determination of an individual's genotype may be used as an indicator of the need for dietary selenium supplementation to inhibit tumour development. Gene therapy can be used to treat tumours by overexpressing the 15 kDa selenoprotein in cells having an abnormally low amount of 15 kDa selenoprotein. The availability of the murine cDNA allows the production of transgenic and knock-out mice useful for assessment of new therapeutic agents.

Sequence 1216 BP; 359 A; 242 C; 275 G; 340 T; 0 other;

alignment_scores: Quality: 751.00 Length: 162
Ratio: 4.845 Gaps: 0
Percent Similarity: 95.679 Percent Identity: 91.975

alignment_block:

US-09-676-718a-1 x AAZ34468 ..

Align seg 1/1 to: AAZ34468 from: 1 to: 1216

1 MetAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
11 ATGGCGGACAGGCGAGGTGGGTGGCTCGGCGACGCTGCGGCTGCGCTT 60
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyValLeuPheS 34
61 GCTGCTGGCGACTGCGTTTCAAGCGGTGCTGCTGCTGGGGCAGAGTTTG 110
34 eSerGluAlaCysArgGlyuLeuGlyPheSerSerAsnLeuLeuCysSer 50
111 GCTCAGAGGCGATGCAGAGAGTTGGTTTCTCCACAACTTCTCTGCACG 160
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
161 TCTTGGCATCTCTTGGAGACAGTTTAATCTGCTCCACCTGCACCTGTTTG 210
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGlyuThrLysLysLeuYra 84
211 CAGAGGGTGTCTGCAGAGAACGACAAATTTGAAACCAAAAGCTGATG 260
84 IaGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
261 CAGAGGCGATCCTTGAAGTCTGCGGATGAAATTTGGGAGTTCCTCA 310
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
311 GTCCAACTTTTTCAGAGATGAACCCAACTTTCAGAGGTCTACA 360
117 nIleLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAsp 134
361 GATCAAGATATGTTGAGGCTCAGACCTGACTAATCACTTTTGGACACA 410
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrPasnThrAspSer 150
411 ACGGGAACATTTGCTGAGAGACTAAGCATTCCTCAATATGGAACACAGCA 460
151 ValGluGluPheLeuSerGlyLysLeuGluArgIle 162
461 GTGGAGAGATTCCTCAGACGAGAACTGGAACGACATA 496

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ43009

seq_documentation_block:

ID AAZ43009 standard; cDNA; 673 BP.

AAZ43009;

01-FEB-2000 (first entry)

Human 5' EST isolated from a cDNA library SEQ ID NO:768.

Human: 5' EST; expressed sequence tag; secreted protein; diagnosis; gene therapy; chromosome mapping; upstream regulatory sequence; forensic; location; development; protein synthesis; stability; regulation; identification; ss.

Homo sapiens.

WO9953051-A2.

21-OCT-1999.

09-APR-1999; 99WO-IB00712.

09-APR-1998; 98US-0057719.

28-APR-1998; 98US-0069047.

PA (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclet A, Giordano J;
 XX WPI: 2000-038446/03.
 XX P-PSDB: AAY65395.
 DR Novel secreted protein 5' expressed sequence tag sequences used in
 XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
 PS Claim 1; Page 570; 837pp; English.
 XX
 CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 XX
 XX Sequence 673 BP; 161 A; 130 C; 193 G; 188 T; 1 other;
 SQ
 alignment_scores:
 Quality: 742.00 Length: 146
 Ratio: 5.153 Gaps: 0
 Percent Similarity: 98.630 Percent Identity: 98.630
 alignment_block:
 US-09-676-718a-1 x AA243009 ..
 Align seg 1/1 to: AA243009 from: 1 to: 673
 1 MetaAlaIaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
 ||||||||||||||||||||||||||||||||||||||||||||
 227 ATGGCGGCTGGGCCGAGGAGGTGTGCTGCGCGCTTTGGGCTACGGTT 276
 17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPheS 34
 ||||||||||||||||||||||||||||||||||||||||||||
 277 GTTGTGGCGGACTGTGTCAAGCGCTGCTTGGTGGGCGAGAGTTT 326
 34 erSerGluAlaCysArgGluLeuGlyPheSerSerAlaLeuLeuGlySer 50
 ||||||||||||||||||||||||||||||||||||||||||||
 327 CATCGGAGGATGAGAGAGATTAGGCTTTTACGCAACTTGGTTGCACG 376
 37 TCTTGTGATCTTCTCGACAGTCAACTGCTTACGTCGAGTCCGATTG 426
 51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
 ||||||||||||||||||||||||||||||||||||||||||||
 377 TCTTGTGATCTTCTCGACAGTCAACTGCTTACGTCGAGTCCGATTG 426
 67 sArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyrA 84
 ||||||||||||||||||||||||||||||||||||||||||||
 427 CAGAGGAGCTGTCTCAGGAGGAAAGCAATTTGAAACCAAAAGCTGTATG 476
 84 lAglyAlaIleLeuGluValCysGly***LysLeuGlyArgPheProGln 100
 ||||||||||||||||||||||||||||||||||||||||||||
 477 CAGGAGCTATTCTTGAAGTTTGATGATAAAATTTGGAAAGGTTCCCTCA 526
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
 ||||||||||||||||||||||||||||||||||||||||||||
 527 GTCCAGAGCTTTTGTAGAGATGATAAACCAAACTTTCAGAGGACTGCA 576

117 nIleLysTyrValArgGlySerAspProValLeuLysLeuAspAspA 134
 ||||||||||||||||||||||||||||||||||||||||||||
 577 AATCAAGTATGTCGTCGTTACAGACCCCTGTATTAAAGCTTTGGACGACA 626
 134 sngLYAsnIleAlaGluGluLeuSerIleLeuLysTirP 146
 ||||||||||||||||||||||||||||||||||||||||||||
 627 ARGGGAACATTGCTGAGAACATGACCATTCCTCAAAATGC 664
 seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA243003
 seq_documentation_block:
 ID AA243003 standard; cDNA: 623 BP.
 XX
 AC AA243003;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST isolated from a cDNA library SEQ ID NO:762.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification; ss.
 XX
 OS Homo sapiens.
 XX
 PN WC0953051-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 09-APR-1999; 99WC-IB00712.
 XX
 PR 09-APR-1998; 98US-0057719.
 PR 28-APR-1998; 98US-0069047.
 XX
 PA (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclet A, Giordano J;
 XX WPI: 2000-038446/03.
 XX P-PSDB: AAY65389.
 DR Novel secreted protein 5' expressed sequence tag sequences used in
 XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
 PS Claim 1; Page 566-567; 837pp; English.
 XX
 CC AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY65438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 623 BP; 143 A; 120 C; 182 G; 175 T; 3 other;
 alignment_scores:
 Quality: 742.00 Length: 146
 Ratio: 5.153 Gaps: 0
 Percent Similarity: 98.630 Percent Identity: 98.630

Quality: 613.50 Length: 147
 Ratio: 4.869 Gaps: 2
 Percent Similarity: 85.714 Percent Identity: 85.034

alignment_block:
 US-09-676-718a-1 x AAA43003 ..

Align seg 1/1 to: AAA43003 from: 1 to: 623

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||
227 ATGGCGGCGTGGCGCGAGTGGTGTCTGGTGGCCGGCTTGGGCTACGGTT 276
  |||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheS 34
  |||
277 GTTGTGGCGGACTGTGCTTCAAGCCGCTGCTTGGGGGCGAGAGTATTT 326
  |||
34 eSerGlyAlaCysArgGlyLeuGlyPheSerSerAsnLeuLeuCysSer 50
  |||
327 CATGGAGGCAATGCAGAGAGTGGCTTTCTAGCAACTTGGCTTTCAGC 376
  |||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
  |||
377 TCTTGATCTTCTCGGACAGTTCACCTGCTCAGCTGATCTGATG 426
  |||
67 sArgGlyCysGlnGlnGlnAlaGlnPheGlyThrLysLysLeuTyrA 84
  |||
427 CAGAGAGATCTGCAGAGAGAACCAATTTGAAACCAAAACCTGATG 476
  |||
84 lAGlyAlaIleLeuGlyValCysGly**LysLeuGlyArgPheProGln 100
  |||
477 CAGGAGCTTCTCTTGAAGTTGTGGATGAAATTTGGAAAGGTCCCT... 523
  |||
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
  ||
524 .....CA 525
  |||
117 nIle.LysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 133
  |||
526 AGTCCAGATATGCTCGTGGTTCAGACCTGTATTAAGCTTTTGGAGCAG 575
  |||
134 AsnGlyAsnIleAlaGlyLeuLeuSerIleLeuLysTyr 146
  |||
576 AATGGACATTTGCTGAGAGACTGAGACATTTCTCAAAATGG 614
  |||
seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA16434
seq_documentation_block:
ID AAA16434 standard; DNA: 578 BP.
XX AAA16434;
AC
XX
XX 14-JUN-2000 (first entry)
DT
XX
DE Human colon cancer differentially expressed nucleotide sequence #439.
XX
XX Colon cancer; detect; differential expression; human; treatment;
XX detect mutation; non-invasive diagnostic method; ds.
OS
XX Homo sapiens.
XX
XX MO200012702-A2.
PN
XX
XX 09-MAR-2000.
PD
XX
XX 30-AUG-1999; 99WO-US19424.
PF
XX
XX 31-AUG-1998; 98US-0098639.
PR
XX 27-JAN-1999; 99US-0117393.
XX
XX (FARB ) BAYER CORP.
PA
XX
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
```

PI Catino TJ, D'ivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 DR WPI; 2000-256641/22.

PT Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer
 XX
 XX
 PS Claim 16; Page 301-302; 345pp; English.

This sequence represents a human nucleotide sequence which is
 differentially expressed in colon cancer cells compared to the expression
 levels in normal cells. The nucleotide sequence can be used as a source
 of primers and probes. The nucleotide sequence is useful for determining
 the phenotype of a cell by detecting the differential expression of the
 sequence relative to a normal cell. The probes derived from the sequence
 can also be used to determine the phenotype of cells in a sample. Probes
 and antibodies which hybridise to the nucleotide sequence can also be
 used to determine the phenotype of a cell. The primers are useful for
 detecting a mutation in a test nucleotide sequence and also for detecting
 cancer, preferably colon cancer. Antibodies against the protein encoded
 by the nucleotide sequence can also be used in a method to detect colon
 cancer. The diagnostic method is non-invasive and accurate for diagnosing
 colon cancer at an early stage.

Sequence 578 BP; 143 A; 102 C; 152 G; 160 T; 21 other;

alignment_scores:

Quality: 552.50 Length: 136
 Ratio: 4.604 Gaps: 3
 Percent Similarity: 88.235 Percent Identity: 86.029

alignment_block:

US-09-676-718a-1 x AAA16434 ..

Align seg 1/1 to: AAA16434 from: 1 to: 578

```

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgIle 17
  |||
43 ATGGCGGCGTGGCGCGAGTGGTGTCTGGTGGCCGGCTTGGGCTACGGTT 92
  |||
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGlnPheS 34
  |||
93 GTTGTGGCGGACTGTGCTTCAAGCGGTGCTTGGGGGCGAGAGTATTT 142
  |||
34 eSerGlyAlaCysArgGlyLeuGlyPheSerSerAsnLeuLeuCysSer 50
  |||
143 CATGGAGGCAATGCAGAGAGTGGCTTTCTAGCAACTTGGCTTTCAGC 192
  |||
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
  |||
193 TCTTGATCTTCTCGGACAGTTCACCTGCTCAGCTGATCTGATG 242
  |||
67 sArgGlyCysGlnGlnGlnAlaGlnPheGlyThrLysLysLeuTyrA 84
  |||
243 CAGAGGATGCTGCAGAGAGAACCAATTTGAAACCAAAACCTGATG 292
  |||
84 lAGlyAlaIleLeuGlyValCysGly**LysLeuGlyArgPheProGln 100
  |||
293 CAGGAGCTATTTCTTGAAGTTTGGAGTGAATTTGGAGAGTTCCTTAA 342
  |||
101 ValGlnAlaPheValArgSerAsp.LysProLysLeuPhe...ArgGlyL 116
  |||
343 GTCCAAACTTTTGTAGACTGATAAACCAAACTTGCAGAGAGCTG 392
  |||
116 euGlnIleLysTyrVal.ArgGlySerAspProValLeuLysLeuLeuAs 132
  |||
393 CCAATTAAGATATGTTNCGTTTCAACCTGATTAAGAGCTTTT 442
  |||
132 pAsp 133
  |||
443 NGAC 446
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL13005
seq_documentation_block:
ID   ABL13005 standard; cDNA; 429 BP.
XX
AC   ABL13005;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 33497.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KM   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-0509231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI: 2001-656860/75.
XX
DR   P-PSDB; ABB68902.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 33497; 21bp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC   sequences (AB101840-AB116175) and the encoded proteins
CC   (ABB5737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 429 BP; 118 A; 103 C; 119 G; 89 T; 0 other;

alignment_scores:
      Quality: 311.00      Length: 114
      Ratio: 3.659      Gaps: 0
Percent Similarity: 74.561      Percent Identity: 51.754

alignment_block:
US-09-676-718A-1 x ABL13005 ..

Align seg 1/1 to: ABL13005 from: 1 to: 429

48 LeucySerSerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAs 64
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1 ATGTTCTCCAGTTCGAAAAAAGCTGGATTTTCGATTTGCAATACCAATCA 50
64 PTrAspCysArgGlyCysCysGlnGluGluAlaGlnPheGluThrTyrSL 81
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
51 GCCTCAGTGAAGCAATGCTGCTGACTTTGGATCAGCAGCGCGCAGCAGC 100
81 ySLeuTYrAlaGlyAlaIleLeuGluValCysGly**LysLeuGlyArg 97
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF63789
seq_documentation_block:
ID   AAF63789 standard; cDNA; 1453 BP.
XX
AC   AAF63789;
XX
DT   03-APR-2001 (first entry)
XX
DE   Human secreted protein gene 1 SEQ ID NO:11.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   neuroprotective; antibacterial; virucide; fungicide; opthalmological;
KW   vulnary; autoimmune disease; hyperproliferative disorder; cancer;
KW   cardiovascular disorder; cerebrovascular disorder; infection;
KW   nervous system disorder; ocular disorder; chemotaxis; food additive;
KW   secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200077021-A1.
XX
PD   21-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15135.
XX
PR   11-JUN-1999; 99US-0138632.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR   WPI: 2001-071257/08.
XX
DR   P-PSDB; AAB75240.
XX
PT   Nucleic acid molecules encoding human secreted proteins, used in
PT   preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT   Parkinson's diseases and cancers -
XX
PS   Claim 1; Page 440; 530pp; English.
XX
CC   This invention relates to polynucleotide sequences AAF63789 - AAF63836
CC   which encode human secreted proteins AAB75260 - AAB75287. Included in the
CC   invention are protein sequences AAB75288 - AAB75341 which are fragments
CC   of the secreted proteins and amino acid sequences with which these
CC   fragments share homology. Examples of the activities of the proteins and
CC   polynucleotides and the activities of their agonists and antagonists
CC   include; immunosuppressive; antiarthritic; antirheumatic;
CC   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
CC   neurotropic; neuroprotective; antibacterial; virucide; fungicide;
CC   opthalmological; and vulnary activity. The protein and polynucleotide
CC   sequences, their agonists and antagonists may be useful for treating,
```

CC preventing and diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders
CC e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiodenias, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Included in the invention are
CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used
CC in the identification and characterisation of the DNA and protein
CC sequences of the invention.

SO Sequence 1453 BP; 473 A; 209 C; 229 G; 542 T; 0 other;

alignment_scores:

Quality: 302.00 Length: 60
Ratio: 5.033 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-676-718a-1 x AAF63789 ..

Align seg 1/1 to: AAF63789 from: 1 to: 1453

103 AlaphValArgSerAspProValLeuLysLeuLeuAspAsnGlu 119
|||||
226 GCTTTGTTAGAGTGTATTAACCAACCTCTCAGAGACTGCATAATCAA 275
119 sTyrValArgGlySerAspProValLeuLysLeuLeuAspAsnGlu 136
|||||
276 GTATGTCCTCGGTTCAGACCTGTATTAAGCTTTTGACAGACATGGGA 325
136 snlleAlaGluGluLeuSerTleLeuLysTrpAsnThrAspSerValGlu 152
|||||
326 ACATTCTCTGAAGAACTGACGATCTCTCAATATGAAACACAGACGTGTGAA 375
153 GluPheLeuSerGluLysLeuGluArgTle 162
|||||
376 GAATTCCTGAGTGAATAAGTTGGAACGACATA 405

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLI3004

seq_documentation_block:

ID ABLI3004 standard; CDNA; 2746 BP.

XX ABLI3004;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 33494.

KM Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PMD, Myers EM;

XX WPI; 2001-656860/75.
DR P-PSDB; ABB688901.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Claim 1; SEQ ID NO 33494; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBI57737-ABBI72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 2746 BP; 853 A; 516 C; 546 G; 831 T; 0 other;

alignment_scores:

Quality: 274.00 Length: 97
Ratio: 3.753 Gaps: 0
Percent Similarity: 75.258 Percent Identity: 53.608

alignment_block:

US-09-676-718a-1 x ABLI3004/rev ..

Align seg 1/1 to reverse of: ABLI3004 from: 1 to: 2746

65 ProAspGlyArgGlyCysGlnGluAlaIlePheGluTrpLys 81
|||||
1378 CTTGAGGTAAAGCAATGCTGCACTTGTGATCAGACGCGCGGACAGCG 1329
81 sLeuTyrAlaGlyAlaIleLeuGluValCysGly**LysLeuGlyArgp 98
: ||||| |||||
1328 GACATATGCCAAGCAATTCGTGAGGTGTGACCTGCMAATTCGGGCGCT 1279
98 heProGluValGlnAlaPheValArgSerAspLysProLysLeuPheArg 114
1278 ATCCGACAGATTGACGCTTATTCAAAGCGCGGACGACCTGCCAATTCGCC 1229
115 GlyLeuGlnIleLysTyrValArgGlySerAspProValLeuLysLeu 131
: |||||
1228 AACCTGCAGATCAATACGTAAAGGAGCTGTGCTGTTAAGCTCCT 1179
131 uAspAspAsnGlyAsnIleAlaGluGluLeuSerTleLeuLysTrpAsnT 148
|||||
1178 AGATGCCAGTGTGCAAAAGTGCAGAGACGTTGTCCATTAACCAAGTGCANCA 1129
148 hrAspSerValGluGluPheLeuSerGluLysLeuGluArg 161
|||||
1128 CAGACACTGTGAGAGAGTCTTCGAAACGCAATCTGGCCCAAG 1088

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA47500

seq_documentation_block:

ID AA47500 standard; DNA; 682 BP.

XX AA47500;

DE 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 54045.

KM Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147203.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

```

PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158332.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

alignment_scores:
      Quality: 226.50      Length: 146
      Ratio: 2.157      Gaps: 4
Percent Similarity: 71.918      Percent Identity: 32.877

alignment_block:
US-09-676-718a-1 x AAC47500 ..

Align seg 1/1 to: AAC47500 from: 1 to: 682

17 LeuLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGluAlaGluPhe 33
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 TTGATATGTCAGATCGACGATCTCAGCAAAAGAG.....CAGCT 78
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
33 eSerSerGluAlaCysArgGluLeuGluPheSerSerAsnLeuLeuLeuYss 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
79 CAGACACAAAGAAATGAGATCTAGAGGTTAGGGCTTGACACTGCTGCT 128
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 eSerSerCysAspLeuLeuGluPhe...AsnLeuLeuGlnLeuAspPro 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
129 CCGATTGCCACCTCAGCTCTGCAATACCTTAAGCAAGCAAGGTTGTAATCT 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 AspCysArgGlyCysGlnGluGluAlaGlnPheGluThrLysLysLe 82
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
179 GATTGCTTGAAGATGCTGCTGATGATTCGAGCATTCATGACCAAGGT 228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
82 u...TyrAlaGluAlaLeuLeuGluValCysGly**LysLeuGlyLysr 98
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 TACCTATTCTGGCGCTATATTATGAGTGTGATGAGGAAGTGTGTTTCT 278

```

```

98 heProGlnValGlnAlaPheValArgSerAspLysProLysLeuPheArg 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
279 ATCCGAAATGTTGCTTTCATTCAGAACAGAAAGAAAG...TTCCCT 325
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
115 GlyLeuGlnIleLysTyrValArgGlySerAspProValLeuLysLeu 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
326 AGCGTCAAGTTCAGTACATTTTCACCTCACCCCAAGTTGATCATGCT 375
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 uAspAspAsnGlyAsnIleLeaGluGluLeuSerIleLeuLysTrpAsn 148
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
376 CGACGAAGACGGTCAACATAAGCAATCATATTAGATGACAACTGGAAC 425
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
148 hrAspSerValGluGluPheLeuSerGluLysLeuGlu 160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
426 GCGAACAATCATCTCAGTATATGCGAGAGAAAGTCAAG 463

seq_name: /SIDSI/gcgdata/mold-geneseq/geneseqn-emb1/NA2000.DAT:AAC34325
seq_documentation_block:
ID AAC34325 standard; DNA; 695 BP.
XX
AC AAC34325;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6246.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200DEP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.

```

PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 23-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144335.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159325.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159329.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159637.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0145086.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145145.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145228.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0161099.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161405.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161406.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161920.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161992.
		PR 28-OCT-1999;	99US-0161993.


```

PR 29-OCT-1999; 99US-0162142.
alignment_scores:
    Quality: 226.50      Length: 146
    Ratio: 2.157        Gaps: 4
Percent Similarity: 71.918    Percent Identity: 32.877
alignment_block:
US-09-676-718A-1 x AAC34325 ..
Align seg 1/1 to: AAC34325 from: 1 to: 695

17 LeuLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPh 33
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 TTGATGTTGGCATCGACGATCTCGACAAAGAG.....CAGCT 91
33 eSerSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuGly 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92 CAGCACAAGAAATGTAGCATCTAGCGGTTTACGGCTCTGCTCTGCT 141
50 eSerCysAspLeuLeuGlyGlnPhe..AsnLeuLeuGlnLeuAspPro 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 CCGATGCCACTCCTCTCGAATACGTTAAGACCAAGATTGCTATCT 191
66 AspCysArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLe 82
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
192 GATTGCTTGAATGTTGTGCTGATGATTCGAGGATTCATGACCAAGT 241
82 u...TyralGlyAlaIleLeuGlnValCysGly**LysLeuGlyLysP 98
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:
242 TACCTATTCGCGCCCTATATTAGAGTGCTGATAGCAAGTTGGTTTCT 291
98 heProGlnValGlnAlaPheValArgSerAspLysProLysLeuPheArg 114
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
292 ATCTGAAATGTTGTTGCTTCATTGAAAGAGAAAGAAAG...TTCCT 338
115 GlyLeuGlnIleLysTyrrValArgGlySerAspProValLeuLysLeu 131
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
339 AGCGTCAAACTCAGTACATTTCACACTCACCCCAAGTTGATCATGCT 388
131 uAspAspAsnGlyAsnIleAlaGlnGluLeuSerIleLeuLysTrpAsnT 148
   |||:|||||:|||||:|||||:|||||:|||||:|||||
389 CGACGAAGACGCTGACATAGCAATATTAAGATGACACACTGGAAC 438
148 hrAspSerValGluGluPheLeuSerGluLysLeuGln 160
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
439 GCGAACATCTACTGCACTATATGCGAGAGAAAGTCAAG 476

seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA57362
seq_documentation_block:
ID ABA57362 standard; DNA; 483 BP.
XX
AC ABA57362;
XX
DE 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #5667.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.

```

```

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 5667; 639bp + sequence listing; English.
XX
33 The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 483 BP; 139 A; 91 C; 105 G; 148 T; 0 other;

alignment_scores:
    Quality: 172.50      Length: 86
    Ratio: 2.875        Gaps: 1
Percent Similarity: 69.767    Percent Identity: 51.163
alignment_block:
US-09-676-718A-1 x ABA57362 ..
Align seg 1/1 to: ABA57362 from: 1 to: 483

1 MetaIaIaGlyProSerGlyCysLeuValProIaPheGlyLysArgLe 17
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
241 GTGCGACAGAGAGAGATGATGCTAGACTGCGCATTTGGATGCAAGTT 290
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyAlaGluPhe 34
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
291 GCCAAGACCACTGATCCCAAGCGCTAGCTGATTTGACACAGTGTAC 340
34 eSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuLysSer 50
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
341 TTCTCTT.....GGCTCCCTTCACCAACTTCTCTGCTGCTGC 374
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
375 TCTTGTGACCTTCTGACGCTTTACCTGTTCAGCTGGATCTTAAGT 424
67 sArgGlyCysCysGlnGluAlaGlnPheGluThrLysLysLeuTyra 84
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
425 CAGAGACTATATCAGAGAAAGTTCAATTGAAACAGAACGCTCGATG 474
84 laGlyAla 86
   : |||:|||||:|||||:|||||:|||||:|||||:|||||
475 TAGGAGCT 482

seq_name: /SIDSL/gcgsdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK05398
seq_documentation_block:
ID AAK05398 standard; DNA; 483 BP.
XX
AC AAK05398;
XX
DE 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 5389.
XX

```



```

|||||
105 CATGGAGGCGATGCGAGAGCTTAGGCTTTTCTACCACTTCTTGGACG 154
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
155 TCTTGATATCTTCTCGGACAGTTCACCTGCTTCACTGATCTGATG 204
67 sArgGlyCysCysGlnGlnGlnPheGlnPheGlnThrlsLeuTyra 84
|||||
205 CAGGAGATGCTGTCAGAGGAGCAATTTGAACCAAAAAGCTGTATG 254
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
255 CAGAGCTATCTTGAAGTTTGTGATGAAATTTGGAGAGTTCCCTCAA 304
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
305 GTCCAGACTTTTGTAGAGATGATAACCCAACTGTTCAAGAGACGCA 354
117 nIleLysTyValArgGlySerAspProValLeuLysLeuAspAspa 134
|||||
355 AATCAAGTATGCTCGTGTGACAGCCCTGTATTAAGCTTTTGACGACA 404
134 snGlyAsnIleAlaGluGlnLeuSerIleLeuLysTyrPasnThrAspSer 150
|||||
405 ATGGGAACATTGCTGAAGAACTGACATTTCTCAATGGACACACAGACAGT 454
151 ValGluGlnPheLeuSerGluLysLeuGlnArgTle 162
|||||
455 GTAGAGAAATCTGAGTGAAGAAAGTTGGAACGACATA 490
seq_name: gb_pr:BC005294

```

```

seq_documentation_block:
LOCUS BC005294 1299 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone MGC:12358 IMAGE:4041321, mRNA, complete cds.
ACCESSION BC005294
VERSION BC005294.1 GI:13529007
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1299)
Strausberg, R.
Direct Submission
Submitted (27-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@pax1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

```

REMARK
COMMENT
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAL Plate: 16 Row: b Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3095110.
FEATURES
source
Location/Qualifiers
1..1299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:12358 IMAGE:4041321"

```

```

/tissue_type="Kidney, hypernephroma"
/clone_id="NIH_MGC_58"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
31..309
/codon_start=1
/product="Unknown (protein for MGC:12358)"
/protein_id="AAH05294.1"
/db_xref="GI:13529008"
/translation="MAGPSGCLVPAFGRLLATVLAQVSAFSAEFSSEACREIGFS
SNLCSGCDLGFNLIDLPDRCGCCQEEAOFETIKLYAGAILVEVC"
BASE COUNT 419 a 223 c 264 g 393 t
ORIGIN

alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765.

alignment_block:
US-09-676-718A-1 x BC005294 ..

Align seg 1/1 to: BC005294 from: 1 to: 1299

1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgLe 17
|||||
31 ATGGCGGCTGGGCCAGAGTGTGCTGTGCGCGCGCTTGGGCGATACGGTT 80
17 uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyValAlaLubes 34
|||||
81 GTTGTGGCGACTGTGCTCAAGCGGTGTCTGCTTTGGGGCAAGATTTT 130
34 erSerGluAlaCysArgGluLeuGlnLysPheSerAsnLeuLeuCysSer 50
|||||
131 CATGGAGGCGATGCGAGAGAGTGTAGGCTTTTCTACCACTTCTTGGACG 180
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
181 TCTTGATGATCTTCGCGACAGTTCACCTGCTCAGCTGAGATCCTGATG 230
67 sArgGlyCysCysGlnGlnGlnAlaGlnPheGluThrLysLysLeuTyra 84
|||||
231 CAGAGGATGCTGTCAAGAGGAAACACATTTGAAACCAAAAAGCTGTATG 280
84 laGlyAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
281 CAGAGGCTATCTTGAAGTTGTGGATGAATAATTTGGAGAGTTCCCTCAA 330
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
331 GTCCAGACTTTTGTAGAGATGAATTAACCCAACTGTTCAAGAGACTGCA 380
117 nIleLysTyValArgGlySerAspProValLeuLysLeuLeuAspAspa 134
|||||
381 AATCAAGATATCTCCGTGTGACACCTGTATTTAAAGCTTTTGAAGACA 430
134 snGlyAsnIleAlaGluGlnLeuSerIleLeuLysTyrPasnThrAspSer 150
|||||
431 ATGGGAACATTGCTGAAGAACTGAGCATTTCAATGGAACACAGACAGT 480
151 ValGluGlnPheLeuSerGluLysLeuGlnArgTle 162
|||||
481 GTAGAGAAATCTGAGTGAAGAAAGTTGGAACGACATA 516
seq_name: gb_pr:AF288991
seq_documentation_block:
LOCUS AF288991 1519 bp mRNA linear PRI 12-NOV-2000
DEFINITION Homo sapiens 15 kDa selenoprotein (SEP15) mRNA, complete cds;
alternatively spliced.
ACCESSION AF288991
VERSION AF288991.1 GI:11138954

```

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 1519)
Kumaraswamy, E., Malikh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y., Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J., Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE Structure-expression relationships of the 15 kDa selenoprotein gene: possible role of the protein in cancer etiology
JOURNAL J Biol Chem. 275 (45), 35540-35547 (2000)
PUBMED 10945981
AUTHORS 2 (bases 1 to 1519)
Kumaraswamy, E., Malikh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y., Kwon, S., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J., Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Department of Biochemistry, University of Nebraska, N151 Beadle Center, Lincoln, NE 68588, USA
FEATURES
source 1..1519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p31"
1..1519
/gene="SEP15"
5..493
/gene="SEP15"
/note="alternatively spliced; long alternative transcript; contains SECIS element; selenocysteine"
/transl_except=(pos:281..283,aa:OTHER)
/codon_start=1
/product="15 kDa selenoprotein"
/protein_id="AAG31556.1"
/db_xref="GI:11138955"
/translation="MAAGPSGCLVPAPFGLRLLATVLAQVAFGEFSSSEACREIGFS
SNLACSSCDLGFQNLQIDPDCRCGCCOEQFETKLYAGAILLEVCCKLGRFOVQ
AFVRSRDKRLFRGLQIKYVRGSDPYLKLDDNGNTAELESLIKNTDSVEPFLSEKLE
RI"
BASE COUNT 458 a 245 c 292 g 524 t
ORIGIN
alignment_scores:
Quality: 819.00 Length: 162
Ratio: 5.119 Gaps: 0
Percent Similarity: 98.765 Percent Identity: 98.765
alignment_block:
US-09-676-718A-1 x AF288991 ..
Align seg 1/1 to: AF288991 from: 1 to: 1519
1 MetAlAlAGlYProSerGlyCysLeuValProAlAPheGlyLysArgLe 17
|||||
5 ATGGCGGGCGGCGCCAGTGGGTCTGTGTCGGCGCTTGGCTACGGTT 54
|||||
17 uleuLeuAlAPhrValLeuGlnAlaValSerAlAPheGlyAlAGluPhes 34
|||||
55 GTTGTGGGCGACATGCTCTCAACGCGTGTCTTTGGGGCGAGATT 104
|||||
34 exSerGluAlaCysArgGluLeuGlyPheSerSerAsnLeuLeuCysSer 50
|||||
105 CATCGAGAGCATGCAGAGACTTTCAGCAACTTTCCTTGCACG 154
|||||
51 SerCysAspLeuLeuGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
|||||
155 TCTTGTGATCTTCTCGACAGTTCACCTGCTCAGCTGATCCTGATG 204
|||||
67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyr 84
|||||

205 CAGAGCATGCTGTACAGAGACACACATTTGAACCAAAAGCTGTATG 254
84 laGlaValAlaLeuGluValCysGly**LysLeuGlyArgPheProGln 100
|||||
255 CAGAGCATGCTGTGAAGTTTGTGGATGAATAATGGAGAGTTCCTCA 304
|||||
101 ValGlnAlAPheValArgSerAspLysProLysLeuPheArgGlyLeuG 117
|||||
305 GTCCAGCTTTTGTAGAGTGAATAAACCAACTGTTCAGAGACTGCA 354
|||||
117 nileLysTyrValArgLysSerAspProValLeuLysLeuLeuAsp 134
|||||
355 AATCAAGTATGTCCTGCTGACACCTGTATTAACCTTTTGACACACA 404
|||||
134 snGlyAsnIleAlaGluGluLeuSerIleLeuLysTyrAsnThrAspSer 150
|||||
405 ATGGAAACATGCTGGAAGAACTGAGCTTCTCAATGGAACACAGACAG 454
|||||
151 ValGluGluPheLeuSerGluLysLeuGluArgLe 162
|||||
455 GTAGAGAAATTCCTGAGTGAAGATTGGAACGACATA 490
seq_name: gb_pr:BC016359
seq_documentation_block:
LOCUS BC016359 1562 bp mRNA linear PRI 05-NOV-2001
DEFINITION Homo sapiens, clone MGC:24512 IMAGE:4096759, mRNA, complete cds.
ACCESSION BC016359
VERSION BC016359.1 GI:16741016
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1562)
Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
Contact: amadanes@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 32 Row: 9 Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
FEATURES
source 1..1562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:24512 IMAGE:4096759"
/tissue_type="Bone marrow, chronic myelogenous leukemia"
/clone_id="NIH_MGC_54"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
18..296
/codon_start=1
/product="Unknown (protein for MGC:24512)."
/protein_id="AAH16359.1"
/db_xref="GI:16741017"

```

/translation="MAAGPSGCVPAFGILRLILATYQANVARGAERSRACBELGS
SNLCSDDLGFQNLPLDIPVRCGCCOEAEQETKLYAGALILEVCG"
BASE COUNT      492 a      246 c      298 g      526 t
ORIGIN

alignment_scores:
    Quality: 819.00      Length: 162
    Ratio: 5.119      Gaps: 0
    Percent Similarity: 98.765      Percent Identity: 98.765

alignment_block:
US-09-676-718A-1 x BC016359

Align seg 1/1 to: BC016359 from: 1 to: 1562

1  Meta1a1a1ag1ProSerG1yCysLeuValProAlaPheG1yLysArg1e 17
|||||
18  ATGGCGGCGTGGGCGGAGTGGGTCTGTGTCGCCGCGCTTTGGGCTACGGTT 67
|||||
17  uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheG1yAlaGluPheS 34
|||||
68  GTTGTGGCGACTGTCTTCAAGCGGTCTCTCTTTGGGGCAGAGTTT 117
|||||
34  eSerG1uAlaCysArg1uLeuG1yPheSerSerAsnLeuLeuCysSer 50
|||||
118  CATCGAGGACATGCAGAGAGTtagGCTTTTCTAGCAACTTGTCTTGACG 167
|||||
51  SerCysAspLeuLeuG1yGlnPheAsnLeuGlnLeuAspProAspCy 67
|||||
168  TCTTGTGATCTTCTTCGACAGTTCACCTGCTTACGTCGATCCTGATTTG 217
|||||
67  sArgG1yCysCysGlnGluGluAlaGlnPheG1uThrLysLysLeuTYR 84
|||||
218  CAGAGAGATGCTGTGCAGAGGAGGACAAATTGCAACCAAAAGCTGATAG 267
|||||
84  1aG1yAla11leuGln1uValCysG1y**LysLeuG1yArpPheProGln 100
|||||
268  CAGGAGCTATTCTTGAAGTTGTGATGAAATTTGGAAAGTTCCCTCA 317
|||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgG1yLeuG1 117
|||||
318  GTCCAAAGCTTTTGTAGAGATGATTAACCAACTGTTCAAGAGACTGCA 367
|||||
117  n11eLysTyrValArgLysSerAspProValLeuLysLeuLeuAspAspA 134
|||||
368  AATCAAGATGTCCTGCTGTCAGACCTGTATTAAGCTTTTGGACGACA 417
|||||
134  sngLysn11lealag1uLeuSer11leuLysTyrAsnThrAspSer 150
|||||
418  ATGGGAACATTGCTGAGAACTGAGCAATTCANATGGAAACAGACAGT 467
|||||
151  ValGluGluPheLeuSerG1uLysLeuG1uArg1le 162
|||||
468  GTAGAGAAATTCCTGATGCAAAAGTTTGGACGCATTA 503

seq_name: gb_10:AF288740

seq_documentation_block:
LOCUS      AF288740      1463 bp      mRNA      linear      ROD 13-NOV-2000
DEFINITION Mus musculus 15 kDa selenoprotein (Sep15) mRNA, complete cds.
ACCESSION  AF288740
VERSION    AF288740.1  GI:11139619
KEYWORDS
SOURCE     house mouse.
           Mus musculus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 1463)
           Kumaraswamy, E., Malykh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
           Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
           Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
           Structure-Expression Relationships of the 15-kDa Selenoprotein

```

```

JOURNAL    Gene. POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY
PUBMED     J. Biol. Chem. 275 (45), 35540-35547 (2000)
REFERENCE  10945981
AUTHORS    2 (bases 1 to 1463)
           Kumaraswamy, E., Malykh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
           Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
           Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
TITLE      Direct Submission
JOURNAL    Submitted (21-JUL-2000) Department of Biochemistry, University of
           Nebraska, N51 Beadle Center, Lincoln, NE 68588, USA
FEATURES   Location/Qualifiers
           source          1..1463
                           /organism="Mus musculus"
                           /db_xref="taxon:10090"
           gene            1..1463
                           /gene="Sep15"
                           /note="Sep15"
                           /note="selenocysteine"
                           /codon_start=1
                           /transl_except="(pos:523..525,aa:OTHER)
                           /product="15 kDa selenoprotein"
                           /protein_id="MAG31765.1"
                           /db_xref="GI:11139620"
           CDS              247..735
                           /translation="MAAGGCGWLRPALGLRLILATYQANVARGAERSRACBELGS
                           SNLCSDDLGFQNLPLDIPVRCGCCOEAEQETKLYAGALILEVCGXKLGRRPVO
                           AFVRSKPKLFRGLQIKYVGRSDPVILKLDNNGNIAEELILKMTNDSVEEFLSEKLE
                           RI"
                           /note="SECTIS element; causes positions 523..525 (TGA) to
                           be translated as selenocysteine"
           misc_feature     1276..1371
                           /gene="Sep15"
                           /note="SECTIS element; causes positions 523..525 (TGA) to
                           be translated as selenocysteine"

BASE COUNT      419 a      299 c      341 g      404 t
ORIGIN

alignment_scores:
    Quality: 747.00      Length: 162
    Ratio: 4.819      Gaps: 0
    Percent Similarity: 95.679      Percent Identity: 91.358

alignment_block:
US-09-676-718A-1 x AF288740

Align seg 1/1 to: AF288740 from: 1 to: 1463

1  Meta1a1a1ag1ProSerG1yCysLeuValProAlaPheG1yLysArg1e 17
|||||
247  ATGGCGGCGAGGCGCAGGTGGGTGCTGCGGCCCTGCTGGGCTGCGCTT 296
|||||
17  uLeuLeuAlaThrValLeuGlnAlaValSerAlaPheG1yAlaGluPheS 34
|||||
297  GCTGTGCGACTGTGCTTCAAGCGGCGTCTCTCTGGGGGAGAGTTT 346
|||||
34  eSerG1uAlaCysArg1uLeuG1yPheSerSerAsnLeuLeuCysSer 50
|||||
347  CGTCAAGGACATGCAGAGAGTGGGTTCTCCAGCAACTTGTCTGACG 396
|||||
51  SerCysAspLeuLeuG1yGlnPheAsnLeuGlnLeuAspProAspCy 67
|||||
397  TCTTGTGATCTTCTTCGACAGTTCATCTGCTCCACATGATGATCTGTTG 446
|||||
67  sArgG1yCysCysGlnGluGluAlaGlnPheG1uThrLysLysLeuTYR 84
|||||
447  CAGAGGCTGTCTGAGGAAGACAAATTGCAACCAAAAGCTGATG 496
|||||
84  1aG1yAla11leuGln1uValCysG1y**LysLeuG1yArpPheProGln 100
|||||
497  CAGGAGCATCTTGAAGTCTGCGATGAAATTTGGGAGGTTCCCTCAA 546
|||||
101  ValGlnAlaPheValArgSerAspLysProLysLeuPheArgG1yLeuG1 117
|||||
547  GTCCAAAGCTTTTGTAGAGATGATTAACCAAACTTCTCAGAGGTCTACA 596
|||||

```

117 ntletyTyValArglySerAspProValLeuLysLeuLysAspAsp 134
 |||||
 597 GATCAAGTATGTCAGAGCTCAGACCTGTACTAAACCTTTTGACACACA 646
 134 snlglyasnilealaglululeuSerlleuLysTrpAsnThrAspSer 150
 |||||
 647 ACGGGAACATGCTGAGAACTAAGCATCTCAAGTGAACACAGACAGT 696
 151 ValGlulupheleuSerGlulysLeuGlulArglle 162
 |||||
 697 GTGGAAGAGTTCCTGAGCGAGAGTGTGAACGACATA 732
 seq_name: gb_ro:BC019792
 seq_documentation_block:
 LOCUS BC019792 1238 bp mRNA linear ROD 22-JAN-2002
 DEFINITION Mus musculus, similar to 15-kDa selenoprotein, clone MGC:30351
 IMAGE:5003585, mRNA, complete cds.
 ACCESSION BC019792.1 GI:18044683
 VERSION BC019792.1 GI:18044683
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCES
 1 (bases 1 to 1238)
 Strausberg, R.
 Direct Submission
 Submitted (19-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.ncl.nih.gov
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amgebcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov
 Series: IRAX Plate: 41 Row: J Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Similarity but not
 identity to protein.
 FEATURES
 source
 1..1238
 location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:30351 IMAGE:5003585"
 /tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month
 old mouse. Taken by biopsy."
 /clone_1lb="NCI_CGAP_Mam2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 15..293
 /codon_start=1
 /product="Similar to 15-kDa selenoprotein"
 /protein_id="AAH19792.1"
 /db_xref="GI:18044684"
 /translation="MAAGOGWLRPALGLRLLIATFAQVAVSALGAEPASACRELGFS
 SNLCSDDLGFQFNLPLDPVRCGCCCEAOEFETKITYAGAILVECG"
 BASE COUNT 369 a 243 c 279 g 347 t
 ORIGIN

alignment_scores:
 Quality: 745.00 Length: 162
 Ratio: 4.806 Gaps: 0
 Percent Similarity: 95.679 Percent Identity: 91.358
 alignment_block:
 US-09-676-718a-1 x BC019792 ..
 Align seq 1/1 to: BC019792 from: 1 to: 1238
 1 MetAlaAlaGlyProSerGlyCysLeuValProAlaPheGlyLysArgle 17
 |||||
 15 ATGGCGGCGAGGCGAGGCTGGGTGGCCGCCCTGCTGGGGCGCCCTT 64
 17 uleuLeuAlaThrValLeuGlnAlaValSerAlaPheGlyLysArgle 34
 |||||
 65 GCTGCTGGCGACTGCTTTCAAGCGGTCTGCTCTGGGGCAGAGTTTG 114
 34 erSerGluAlaCysArgGluLeuGlnPheSerSerAsnLeuLysCysSer 50
 ::|||
 115 GCTCAGAGGCGATGAGAGAGTGGGTTCTCCACCACTGCTCTGCAGC 164
 51 SerCysAspLeuLeuGlnPheAsnLeuLysLeuLysProAspCys 67
 |||||
 165 TCTTGCATCTCTTGGACAGATTAACTGCTCCACCTGCCTGTTG 214
 67 sArgGlyCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyra 84
 |||||
 215 CAGAGGCTGCTGTCAGAGAGCAACCAATTGAAACCAAAAGCTGATG 264
 84 laGlyAlaIleLeuGlnValCysGly**LysLeuGlyArgPheProGln 100
 |||||
 265 CAGAGCCATCTTGAAGTCTGCGAGTGAATAATGGGAGGTGCTCA 314
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeu 117
 |||||
 315 GTCCACCTTTTGTGAGAGATGAATCAACCTTCAAGAGTCTCA 364
 117 ntletyTyValArglySerAspProValLeuLysLeuLysAspAsp 134
 |||||
 365 GATCAAGTATGTCAGAGCTCAGACCTGTACTAAACCTTTTGACACACA 414
 134 snlglyasnilealaglululeuSerlleuLysTrpAsnThrAspSer 150
 |||||
 415 ACGGGAACATGCTGAGAACTAAGCATCTCAATGTAAGACACAGACAGT 464
 151 ValGlulupheleuSerGlulysLeuGlulArglle 162
 |||||
 465 GTGGAAGAGTTCCTGAGCGAGAGTGTGAACGACATA 500
 seq_name: gb_ro:AF390544
 seq_documentation_block:
 LOCUS AF390544 1236 bp mRNA linear ROD 16-JUL-2001
 DEFINITION Rattus norvegicus 15 kDa cytosolic selenoprotein mRNA, complete
 cds.
 ACCESSION AF390544
 VERSION AF390544.1 GI:14719271
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCES
 1 (bases 1 to 1236)
 Roethlein, D., Kyriakopoulos, A. and Behne, D.
 A 15 kDa-selenoprotein in several tissues of the rat
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1236)
 Roethlein, D., Kyriakopoulos, A. and Behne, D.
 Direct Submission
 TITLE

JOURNAL Submitted (11-JUN-2001) SF 6, Hahn-Meitner-Institut, Glienicke
 FEATURES Str. 100, Berlin D-14109, Germany
 Location/Qualifiers
 source
 1..1236
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 1..489
 /note="selenocysteine"
 /codon_start=1
 /transl_except="(pos:277..279,aa:OTHER)
 /product="15 kDa cytosolic selenoprotein"
 /protein_id="AAK73100.1"
 /db_xref="GI:14719272"
 /translation="MAAGGCGWLRPALGLRLATATQAVSALCAERSSRACRELGTS
 SNLSSCDLLGDFNLDPDPCGCGCCQEEAQFTKRLVAGALLVEGAKLRFPQVQ
 AFVRSKPKLFRRGIKRYKRGSDPVKLRLDNGNIABELSLKKXNDSEVEFLSEKLE
 R"
 misc_feature 1085..1132
 /note="selenocysteine insertion sequence (SECIS)"
 polyA_signal 1211..1216
 BASE COUNT 336 a 263 c 287 g 347 t 3 others
 ORIGIN
 alignment_scores:
 Quality: 741.00 Length: 162
 Ratio: 4.812 Gaps: 0
 Percent similarity: 95.062 Percent identity: 91.975
 alignment_block:
 US-09-676-718a-1 x AF390544 ..
 Align seg 1/1 to: AF390544 from: 1 to: 1236
 1 MetalaalaglProserglyCysleuValProalapheglyLysArgpe 17
 |||||
 1 ARGGGGGGCGAGGCGGCGGTGCTCCGCCGCCGCTGGCGCTTGGCTT 50
 17 uleuLeuAlaThrValleuGlnAlaValSerAlaPheGlyAlaGluPheS 34
 |||||
 51 ACTGTGGGAGCTGGCTTCAAGCGGTGTCTCTCTGGGGCGAGAGTTCT 100
 34 eSerGlnAlaIacysArgGluLeuGlyPheSerSerAsnLeuLeuLeuSer 50
 |||||
 101 COTCAGAGCATGCCGGAGGTGGCTTCTCCAGCAACTGCTGTGCACG 150
 51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
 |||||
 151 TCCGTGCAATCCTCTTGACAGCTTAACGTCTTCCACTGTGATCCTGTG 200
 67 sArgGlyCysCysGlnGluGlnAlaGlnPheGluThrLysLysLeuTyrA 84
 |||||
 201 CAGAGGCTGCTGTCAAGAGAGCGCAGTTTAAACCAAAAAAGCTGTATG 250
 84 IagLAlaIleleuGlnuValCysGly**LysleuGlnYArghPheProGln 100
 |||||
 251 CAGGAGCCATCTTGAAGTCTGTGATGAAATGTGGGGAGTTCCCTCA 300
 101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgGlyLeuGln 117
 |||||
 301 GTCCAAAGCTTTTGTCAAGAGCATTAACCAAACTGTTCAAGAGCTTACA 350
 117 nIleLysTyrValArgGlySerAspProValLeuLysLeuLeuAspAspA 134
 |||||
 351 GATCAAGATATGTTCAAGGCTCAGACCTGTACTAAAGCTTTTGGAGACA 400
 134 sngLYAsnIleAlaGlnGluLeuSerIleLeuLysTyrAsnThrAspSer 150
 |||||
 401 ACGGGAACATTGCTGAAGAGCTCAGACATCTCAAGTDDAACACAGACAGT 450
 151 ValGlnGluPheLeuSerGlyLysLeuGlnArgIle 162
 |||||
 451 GTGGAAGAGTTCTCAGCGAGAGAGCTGAGACACATA 486

seq_name: gb:BC021697
 seq_documentation_block:
 LOCUS BC021697 1249 bp mRNA linear PRI 18-JAN-2002
 DEFINITION Homo sapiens, clone MGC:24367 IMAGE:4051151, mRNA, complete cds.
 ACCESSION BC021697
 VERSION BC021697.1 GI:18203835
 KEYWORDS
 SOURCE MGC.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1249)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 NIH-MGC Project help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadasystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 31 Row: 0 Column: 24
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.
 location/Qualifiers
 source
 1..1249
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:24367 IMAGE:4051151"
 /tissue_type="kidney, hypernephroma"
 /clone_lib="NIH_MGC_58"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 37..315
 /codon_start=1
 /product="Unknown (protein for MGC:24367)"
 /protein_id="AAH21697.1"
 /db_xref="GI:18203836"
 /translation="MAAGPSGLVPARGRLRLATVLAQVARSFAERSSRACRELGTS
 SNLSSCDLLGDFNLDPDPCGCGCCQEEAQFTKRLVAGALLVEGAKLRFPQVQ"
 BASE COUNT 400 a 212 c 258 g 379 t
 ORIGIN
 alignment_scores:
 Quality: 701.50 Length: 163
 Ratio: 4.906 Gaps: 2
 Percent similarity: 87.730 Percent identity: 87.117
 alignment_block:
 US-09-676-718a-1 x BC021697 ..
 Align seg 1/1 to: BC021697 from: 1 to: 1249
 1 MetalaalaglProserglyCysleuValProalapheglyLysArgpe 17
 |||||
 37 ARGGGGGGCGAGGCGGCGGTGCTGCTGGCGGCGGCTTGGCGCTACGTT 86
 17 uleuLeuAlaThrValleuGlnAlaValSerAlaPheGlyAlaGluPheS 34


```

|||||
87 GTGTGGGCGACTGCTTCACGGGCTGCTTTGGGGCAGAGTTT 136
34 erserGluaIaCysArgLulLeuGlyPheSerSerAsnLeuLeuCySer 50
137 CATCGGAGGATGACAGAGACTTAGGCTTTCTAGCAACTGCTTTGCAGC 186
51 SerCysAspLeuLeuGlyGlnPheAsnLeuLeuGlnLeuAspProAspCy 67
187 TCTTGGATCTCTCGACAGTTCACCTGCTCAGCTGATGCTCCTGATG 236
67 sArgLylCysCysGlnGlnGlnLulAaGlnPheGlnThrlLysLysLeuTyra 84
237 CAGAGATGCTGTGAGAGAGACACCAATTGAAACCAAAAGCTGATG 286
84 laGlAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGln 100
287 CAGGACTATCTTGAAGTTGTGGATGAAATTTGGAGAGTTCCT... 333
101 ValGlnAlaPheValArgSerAspLysProLysLeuPheArgLysLeuGln 117
334 .....CA 335
117 ntle.LysTyrValArgGlySerAspProValLeuLysLeuLeuAspAsp 133
336 AGTCAGATATGTCGCTGTCAGACCTGTATTAAAGCTTTGGAGAGAC 385
134 AsnGlyAsnIleAlaGlnGluLeuSerIleLeuLysTyrPasnThrAspSe 150
386 AATGGAACTTCTGAGAACTGAGACATCTCAATGAAACACAGACAG 435
150 rValGlnGluPheLeuSerGlnLysLeuGlnArgIle 162
436 TGTAGAAAGATTCTGAGTGAAGTTGGAACCATTA 472
seq_name: gb_ro:BC010481

seq_documentation_block:
LOCUS BC010481 1020 bp mRNA linear ROD 12-JUL-2001
DEFINITION Mus musculus, clone IMAGE:3482046, mRNA, partial cds.
ACCESSION BC010481
VERSION BC010481.1 GI:14714679
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalob@bcm.tmc.edu.
Villalob, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Jellford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: RAK Plate: 6 Row: b Column: 4
This clone was selected for full length sequencing because it

```

```

FEATURES
source
Location/Qualifiers
1..1020
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="3482046"
/tissue_type="Mammary tumor. MAP-TGF alpha model. 7 months
old, gross tissue."
/clone_id="NCI CGAP_Mams"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
<1..181
/codon_start=2
/product="Unknown (protein for IMAGE:3482046)."
/protein_id="AAH10481.1"
/db_xref="GI:14714680"
/translation="GAVRKRLKRSCKQEPSLKSADENMGSLKSLSEVINPNS
SEVRSMEVOTLY"

BASE COUNT 340 a 188 c 199 g 293 t
ORIGIN

alignment_scores:
Quality: 476.00 Length: 94
Ratio: 5.118 Gaps: 0
Percent Similarity: 98.936 Percent Identity: 98.936

alignment_block:
US-09-676-718a-1 x BC010481 ..

Align seq 1/1 to: BC010481 from: 1 to: 1020
69 GTCyCysGlnGlnGlnGlnPheGlnThrlLysLysLeuTyraIaG 85
1 GGGTGGCTGTGAGAGAGAACCAATTGAAACCAAAAGCTGTATGCAGG 50
85 yAlaIleLeuGluValCysGly**LysLeuGlyArgPheProGlnValG 102
51 AGCCATCTTGAAGTGTGCGGATGAATAATGGGAGGTTCCCTCAAGTCC 100
102 lnaIaPheValArgSerAspLysProLysLeuPheArgLysLeuGlnIle 118
101 AAGCTTTGTGAGAGATGAATAAACCAACTCTTCAAGGCTTCAAGATC 150
119 LysTyrValArgGlySerAspProValLeuLysLeuLeuAspAspAsnG 135
151 AAGTATGTTGAGGCTCAGACCTGTACTTAAAGCTTTGGACGACACGG 200
135 yAsnIleAlaGlnGluLeuSerIleLeuLysTyrPasnThrAspSerValG 152
201 GAACATTTGTGAGAGAACTAAGCATCTCAATGAAACAGACAGAGCTGTGG 250
152 lGluPheLeuSerGlnLysLeuGlnArgIle 162
251 AAGATTTCTGAGGAGAGAGATTTGAAACGATTA 282
seq_name: gb_pat:AX330345

seq_documentation_block:
LOCUS AX330345 313 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 854 from Patent W00194629.
ACCESSION AX330345
VERSION AX330345.1 GI:18103323
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Harrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature

```

gene sets
JOURNAL Patent: WO 0194629-A 854 13-DEC-2001;
FEATURES Avalon Pharmaceuticals (US)
source Location/Qualifiers
1. .313
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 64 a 61 c 92 g 96 t
ORIGIN

alignment_scores:
Quality: 392.00 Length: 103
Ratio: 4.455 Gaps: 2
Percent Similarity: 85.437 Percent Identity: 84.466

alignment_block:
US-09-676-718A-1 x AX330345 ..

Align seg 1/1 to: AX330345 from: 1 to: 313

9 LeuValProAlaPheGlyLysArgLeuLeuAlaThrValLeuGlnAl 25
|||||
1 CTGGTCCCGCGCTTGGGCTACGTTGTTGGGACATGCTTCAAGC 50
25 aValSerAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuG 42
|:::|
51 GTTGTCTCTTGGGCGAGATTTCATCGAGGATCGACAGAGATTAG 100
42 LYPheSerSerAsnLeuLeuCysSerSerCysAspLeuLeuGluGlnPhe 58
|||||
101 GCTTTTACGACACTTCTTGCACACTCTTGATCTTCTCGGACAGATTTC 150
59 AsnLeuLeuGlnLeuAspProAspCysArgGlyCysGlnGluAla 75
|||||
151 AACCTGCTTACGCTGATCGATTCAGAGATGCTGTCAGAGAGAGGC 200
75 aGluPheGluThrLysLysLeuTyraGlyAlaIleLeuGluValCysG 92
|||||
201 ACAATTGAAACCAAAAGCTGATGAGAGCTATTCTTGAGTTTGTG. 248
92 Lys**LysLeuGlyArgPhe.....Pro 99
|||||
249 ..GATTAATTGGGAAGTTCCTCAGTCACGCTTGTAGGGTGATTAACA 296
100 GlnValGln 102
|||||
297 ACTGTTTCAG 305

seq_name: gb_in:AF457547

seq_documentation_block:
LOCUS AF457547 624 bp mRNA linear INV 29-JAN-2002
DEFINITION Anopheles gambiae selenoprotein mRNA, partial cds.
ACCESSION AF457547
VERSION AF457547.1 GI:18389880
KEYWORDS
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anophelinae;
1 (bases 1 to 624)
Francischetti,I.M., Valenzuela,J.G. and Ribeiro,J.M.
Towards a catalog for genes and proteins from the salivary gland of
the malaria vector, Anopheles gambiae
Unpublished
JOURNAL 2 (bases 1 to 624)
REFERENCE Francischetti,I.M., Valenzuela,J.G. and Ribeiro,J.M.
AUTHORS Direct Submission
TITLE Submitted (12-DEC-2001) Section of Medical Entomology/Laboratory of
JOURNAL Parasitic Diseases/NIHID, National Institutes of Health, 4 Center
Drive, Building 4, Room 126, MSC 0425, Bethesda, MD 20892-0425, USA

FEATURES Location/Qualifiers
source 1. .624
/organism="Anopheles gambiae"
/db_xref="taxon:7165"
CDS <1.492
/note="salivary protein"
/codon_start=1
/product="selenoprotein"
/protein_id="FAL68777.1"
/db_xref="GI:18389881"
SDYGLIEKHEKCECKDTEADSKLVYPAVALVCTCKFGAVPQIQAFTKSDRPK
FPMITIKYVAGLDPIYALMDQGVKEETISINMWNDFVQEFETRLAKVEDDYIKT
NRV"
BASE COUNT 213 a 118 c 148 g 145 t
ORIGIN

alignment_scores:
Quality: 383.00 Length: 151
Ratio: 3.165 Gaps: 3
Percent Similarity: 80.132 Percent Identity: 48.344

alignment_block:
US-09-676-718A-1 x AF457547 ..

Align seg 1/1 to: AF457547 from: 1 to: 624

16 ArgLeuLeuLeuAlaThrValLeu.....GlnAlaValSerAlaPheG 30
|||||
4 CACCTCTTGGCCATTACGTCGTCTTATTTCGATGATGACGTGATTGG 53
30 YAlaGluPheSerSerGluAlaCysArgGluLeuGly..PheSerSera 46
|||||
54 TGCACAAATTCTCCGCCGAAGATGCGGGAGCTGGGCTTATCAATGCG 103
46 snLeuLeuCysSerSerCysAspLeuLeuGlyGlnPheAsnLeuLeuGln 62
|||||
104 AACTGTTTGTTCGGCGGTGTAGTACTTAAGCATTAAGATGATGATCA 153
63 LeuAspProAspCysArgGlyCysGlnGluGlnAlaGlnPheGluTh 79
|||||
154 CTGAAGAAGACATTGCTTGAATGTTGCCAAAGACACTGAAGCCGATTTC 203
79 Lys...LysLeuTyraGlyAlaIleLeuGluValCysGly**LysL 95
|||||
204 AAAACTGAAGGTGATCCGGCGGAGTGTGAGAGGTGACAGCTCAAGT 253
95 euGlyArgPheProGlnValGlnAlaPheValArgSerAspLysProLys 111
|||||
254 TTGGCGCATATCCACAGATTCAGGGCTTCAATTAAGGACCGACCGGCC 303
112 LeuPheArgGlyLeuGlnIleLysTyraValArgGlySerAspProVal 128
|||||
304 AAGTTTCCCAACCTAACCATCAAGTACGTGCTGTTGGATCCATCGT 353
128 uLysLeuLeuAspAspAsnGlyAsnIleAlaGluGluLeuSerIleLeu 145
|||||
354 GAAGCTAATGCGACGACGAGCGACACTGTAGAAGAAAGCTTTCATTAACA 403
145 yStrpAsnThrAspSerValGluGluPheLeuSerGluLysLeuGluArg 161
|||||
404 AATGGACACCGACACGAGGTTCGAGAGTTTTCGAAACAGCTTTGGCGAAA 453
162 Ile 162
|||
454 GTT 456

seq_name: gb_pr:AF288992

seq_documentation_block:
LOCUS AF288992 64165 bp DNA linear PRI 12-NOV-2000
DEFINITION Homo sapiens 15 kDa selenoprotein (SEPI5) gene, complete cds.

ACCESSION AF288992
 VERSION AF288992.1 GI:11138956
 KEYWORDS
 SOURCE
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 64165)
 Kumaraswamy, E., Maljkh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
 Kwon, S.Y., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Structure-Expression Relationships of the 15-kDa Selenoprotein
 Gene. POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY
 J. Biol. Chem. 275 (45), 35540-35547 (2000)
 PUBMED 10945381
 TITLE 2 (bases 1 to 64165)
 JOURNAL Kumaraswamy, E., Maljkh, A., Korotkov, K.V., Kozayvkin, S., Hu, Y.,
 Kwon, S., Moustafa, M.E., Carlson, B.A., Berry, M.J., Lee, B.J.,
 Hatfield, D.L., Diamond, A.M. and Gladyshev, V.N.
 Direct Submision
 JOURNAL Submitted (24-JUL-2000) Department of Biochemistry, University of
 Nebraska, N151 Beadle Center, Lincoln, NE 68588, USA
 FEATURES
 Source
 1..64165
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1p31"
 join(6777..6873,17484..17651,40213..40276,52814..52863,
 57310..58192)
 /gene="SEPI5"
 /product="15 kDa selenoprotein"
 6777..6798
 /gene="SEPI5"
 6777..58192
 /gene="SEPI5"
 6777..6873
 /gene="SEPI5"
 /number=1
 join(6799..6873,17484..17651,40213..40276,52814..52863,
 57310..57441)
 /gene="SEPI5"
 /note="contains SECIS element; selenocysteine"
 /codon_start=1
 /transl_except=(pos:40246..40248,aa:OTHER)
 /product="15 kDa selenoprotein"
 /protein_id="AAG31557.1"
 /db_xref="GI:11138957"
 /translation="MAAGPSGCLVPAPFGRLRLATVLAQVSAFGAFSSSEACRELGFSS
 SNLACSSCDLIGQENLQLDPCRCGCCCEAOFETKLYAGAILLEVCGXKLGKRPQVO
 AFVRSRDKPLFRGLQIKYVRGSDPYLKLDDNGNIAEELSLIKWNTDSVEEFLSEKLE
 RI"
 17484..17651
 /gene="SEPI5"
 /number=2
 17652..40212
 /gene="SEPI5"
 /note="contains differences with sequence deposited in
 Genbank Accession Number AL121989"
 /number=2
 40213..40276
 /gene="SEPI5"
 /number=3
 52814..52863
 /gene="SEPI5"
 /number=4
 57310..58192
 /gene="SEPI5"
 /number=5
 57442..58192
 /gene="SEPI5"
 3' UTR
 57442..58192
 /gene="SEPI5"
 BASE COUNT 18805 a 12503 c 12005 g 20852 t
 ORIGIN

alignment_scores: Quality: 304.00 Length: 135
 Ratio: 3.102 Gaps: 6
 Percent Similarity: 72.593 Percent Identity: 54.815
 alignment_block:
 US-09-676-718a-1 x AF288992 ..
 Align seg 1/1 to: AF288992 from: 1 to: 64165
 26 valSerAlaPheGlyAlaGluPheSerSergLualAcysArgLueuGI 42
 17484 GTGCTGCTTTGGGCGAGAGTTTTCATCGAGAGCATGACAGAGTTAG 17533
 42 yPheSerSerAsnLueuLueuLueuLueuLueuLueuLueuLueu 59
 17534 CTTTCTTACGACACTGCTTGGCAGCTCTTGATCTTCTGACAGCTTCA 17583
 59 snLueuGlnLueuAspProAspCysArgGlyCysGlnGluLual 75
 17584 ACCTGCTTCAGCTGATCCTGATTCAGAGAGATGCTGAGAGAGCA 17633
 76 GlnPheGlnThrLysLysLysLysLysLysLysLysLysLysLys 92
 17634 CAATTGCAACCAAAAGGAGTCTGTTCTAGTATA..... 17669
 92 y***LysLueuGlyArgPheProGlnValGlnAlaPheValArgSer 108
 17670TTCCCTCCATTTGATGTTAGTAAAGCAAGGT 17703
 109 ..LysProLysLueuPheArgGlyLueuGlnLysLysValArgGlySer 124
 17704 AAAAATGTAAGTTCCTTAAT.....CAATTAAAGTAAAAAGCTCT 17747
 125 AspProValLueuLysLueuAspAspAsnGlyAsnLleAlaGluLue 141
 17748AAGCTTTTA.....CATTTTAAAGTAAAGT 17773
 141 user...LleLueuLysTrpAsnThrAspSerValGluGluPheLueuSerg 157
 17774 CAACGTATTTCTTGTCTGCGCTATGATTCATCAACCAAAATATGAAAG 17823
 157 Luls 158
 17824 AGAAG 17828
 seq_name: gp_htc:AL355981
 seq_documentation_block:
 LOCUS AL355981 155871 bp DNA linear HTG 23-JAN-2001
 DEFINITION Homo sapiens chromosome 1 clone RP11-277G24, *** SEQUENCING IN
 PROGRESS ***, 15 unordered pieces.
 ACCESSION AL355981
 VERSION AL355981.3 GI:9797397
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 155871)
 AUTHORS Plumb, B.
 TITLE Direct Submision
 JOURNAL Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 12, 2000 this sequence version replaced gi:92133687.
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

```

----- Project Information
Center project name: BA277G24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 146127 bases at least Q40
Consensus quality: 149488 bases at least Q30
Consensus quality: 151696 bases at least Q20
Insert size: 154471; sum-of-contigs
Insert size: 176463; 14.5% error; agarose-fp
Quality coverage: 3.58x in Q20 bases; sum-of-contigs Quality
coverage: 3.19x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 15247: contig of 15247 bp in length
15248 15347: gap of 100 bp
15348 21318: contig of 5971 bp in length
21319 21418: gap of 100 bp
21419 25299: contig of 3881 bp in length
25300 25399: gap of 100 bp
25400 30376: contig of 4977 bp in length
30377 30476: gap of 100 bp
30477 33263: contig of 2787 bp in length
33264 33363: gap of 100 bp
33364 72796: contig of 39433 bp in length
72797 72896: gap of 100 bp
72897 98426: contig of 25530 bp in length
98427 98526: gap of 100 bp
98527 106876: contig of 8350 bp in length
106877 106976: gap of 100 bp
106977 111928: contig of 4952 bp in length
111929 112028: gap of 100 bp
112029 119408: contig of 7380 bp in length
119409 119508: gap of 100 bp
119509 137994: contig of 18486 bp in length
137995 138094: gap of 100 bp
138095 142132: contig of 4038 bp in length
142133 142332: gap of 100 bp
142333 144266: contig of 2034 bp in length
144267 144366: gap of 100 bp
144367 148940: contig of 4574 bp in length
148941 149040: gap of 100 bp
149041 155871: contig of 6831 bp in length.
Location/Qualifiers
1. 155871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_1fb="RP11-277G24"
/clone_1lb="RP11-11.1"
1. 15247
/note="assembly_fragment:00782"
fragment_chain:1"
15348. 21318
/note="assembly_fragment:01104"
fragment_chain:1"
21419. 25299
/note="assembly_fragment:00085"
25400. 30376
/note="assembly_fragment:00125"
30477. 33263
/note="assembly_fragment:00175"
33364. 72796
/note="assembly_fragment:00232"
72897. 98426

```

```

misc_feature /note="assembly_fragment:00364"
98527. 106876
misc_feature /note="assembly_fragment:00437"
106977. 111928
misc_feature /note="assembly_fragment:00459"
112029. 119408
misc_feature /note="assembly_fragment:00503"
119509. 137994
misc_feature /note="assembly_fragment:00572"
138095. 142132
misc_feature /note="assembly_fragment:00752"
142233. 144266
misc_feature /note="assembly_fragment:01162"
144367. 148940
misc_feature /note="assembly_fragment:01243"
149041. 155871
misc_feature /note="assembly_fragment:01367"
BASE COUNT 44562 a 31293 c 3116 g 47497 t 1403 others
ORIGIN
alignment_scores:
Quality: 304.00 Length: 135
Ratio: 3.102 Gaps: 6
Percent Similarity: 72.593 Percent Identity: 54.815
alignment_block:
US-09-676-718A-1 x AL355981 ..
Align seg 1/1 to: AL355981 from: 1 to: 155871
26 valserAlaPheGlyAlaGluPheSerSerGluAlaCysArgGluLeuG1 42
|||||GTCCTGCTTTGGGGCAGAGTTCATCGCGCATCGAGATCGATTG 38819
42 yPheSerSerAlaLeuLeuCysSerSerCysAspLeuGluGlnPheA 59
|||||CTTTCTGACCACTGCTTTCGACGCTCTTGATCTTTCGACAGCTTCA 38869
38870 ACCTGCTTCAGCTGATGCTGATTCAGAGATGCTGTCAGAGAGAACCA 38919
76 GlnPheGluThrLysLysLeuTyraAlaGlyAlaLeuGluValCysG1 92
|||||CAATTGGAACCAAAAGCTACTGTTTCTAGTATA..... 38955
38920 y**LysLeuGlyArgPheProGlnValGlnAlaPheValArgSerAsp. 108
|||||TTCCTTCATTTGATGATGTTTATGTAGACAAAGCT 38989
38956 .....LysProLysLeuPheArgGlyLeuGlnIleLysTyraValArgLysSer 124
|||||AAAAATGTAAGTGTCTTAAT.....CAATTAACTGTAACCAACCCCT 39033
125 AspProValLeuLysLeuLeuAspAspAsnGlyAlaIleAlaGluGlu 141
|||||AACCTTTTA.....CAATTGTAAGTAAACT 39059
39034 user...IleLeuLysTrpAsnThrAspSerValGluGluPheLeuSerg 157
|||||CAACGATTTCTGCTGCTGATTCATTAGTCAATCAACCAATATGAAGG 39109
157 Lulys 158
|||||
39110 AGAAG 39114
seq_name: gb_DR:AF267982S2
seq_documentation_block:
LOCUS AF267982S2 1743 bp DNA linear PRI 27-JUN-2000
DEFINITION Homo sapiens 15 kda selenoprotein gene, exon 2.
ACCESSION AF267983

```

```

VERSION      AF267983.1  GI:87455559
KEYWORDS
SEGMENT      2 of 5
SOURCE       human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1743)
AUTHORS      Ryu, M. and Moon, E.
TITLE        The human 15-kDa selenoprotein gene : Characterisation of the
              genomic structure and functional analysis of the promoter
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1743)
AUTHORS      Ryu, M. and Moon, E.
TITLE        Direct Submission
JOURNAL      Submitted (16-MAY-2000) Biological Science, Ajou University, San-5,
              Wonchun-dong, Paidal-gu, Suwon, Kyungki-do 442-749, South Korea
FEATURES
  source      1..1743
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="2075D19"
  exon        795..962
              /number=2
  BASE COUNT  534 a 312 c 342 g 540 t 15 others
  ORIGIN
26 valSerAlaPheGluAlaGluPheSerSerGluAlaCysArgGluLeu 42
42 yPheSerSerAsnLeuLeuCysSerSerCysAspLeuLeuGluPhe 59
845 CTTTCTTACCACTTCTTGGCAGCTTGTGATCTTCTGCGACAGTTCA 894
59 snLeuLeuGlnLeuAspProAspCysArgGlyCysCysGlnGluAla 75
895 ACCTGCTTCAGCTGATCTCGATTGCAAGAGATGCTGTCAGAGAGCA 944
76 GlnPheGluThrLysAllys...LeuTyAlaGlyAlaIleLeuGluVal 91
945 CATTGGAAACCAAAAGGACTGTTTCTAGATATATCTTCATTGGA 994
91 scLy 92
995 TGGT 998
seq_name: gb_pr:HSDJ604K5
seq_documentation_block:
LOCUS      HSDJ604K5 127201 bp DNA linear PRI 24-NOV-2000
DEFINITION Human DNA sequence from clone RP4-604K5 on chromosome 1p22.2-31.1.
              Contains a gene for a 15 kDa selenoprotein and a part of the
              heparan sulfate 2-sulfotransferase gene (K1RA0448). Contains ESTs,
              STSs, GSSs and Cpg islands, complete sequence.
ACCESSION   ALI21989.12 GI:8919105
VERSION     ALI21989.12
KEYWORDS    HTG: Cpg island; K1RA0448; selenoprotein; sulfotransferase.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE    1 (bases 1 to 127201)
AUTHORS      Pearce, A.
TITLE        Direct Submission
JOURNAL      Submitted (17-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
              requests: clonerequests@sanger.ac.uk
COMMENT      On Jul 5, 2000 this sequence version replaced gi:8894634.
              During sequence assembly data is compared from overlapping clones.
              Where differences are found these are annotated as variations
              together with a note of the overlapping clone name. Note that the
              variation annotation may not be found in the sequence submission
              corresponding to the overlapping clone, as we submit sequences with
              only a small overlap as described above.
              This sequence has been finished according to sequence map criteria
              as follows. An attempt is made to resolve all sequencing problems,
              such as compressions and repeats, but not necessarily within known
              annotated human repeat sequence elements (e.g. Alu). Where the
              sequence is ambiguous, there is an annotation using the 'unsure'
              feature key.
              The following abbreviations are used to associate primary accession
              numbers given in the feature table with their source databases:
              Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
              on the WORMPEP database can be found at
              http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
              was generated from part of bacterial clone configs of human
              chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
              Group. Further information can be found at
              http://www.sanger.ac.uk/HGP/Chr1
              RP4-604K5 is from the library RPCI-4 constructed at the Roswell
              Park Cancer Institute by the group of Pieter de Jong. For further
              details see http://bacpac.med.buffalo.edu/
              VECTOR: pCYPAC2
              This sequence is the entire insert of clone RP4-604K5.
FEATURES
  source      1..127201
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="1"
              /map="p22.2-31.1"
              /clone="RP4-604K5"
              /clone_id="RPCI-4"
  repeat_region 1..381
              /note="MLTID repeat: matches 113..504 of consensus"
              /note="74..804
              /note="LM1M4 repeat: matches 7364..7597 of consensus"
              /note="843..925
              /note="LM1M4 repeat: matches 7888..7977 of consensus"
              /note="874..1348
              /note="match: GSS: Em:A0432443"
              /note="match: GSS: Em:A0432443"
              /note="1621..1921
              /note="Alu repeat: matches 1..299 of consensus"
              /note="1951..1982
              /note="MIR repeat: matches 117..148 of consensus"
              /note="2021..2070
              /note="Alu repeat: matches 253..302 of consensus"
              /note="2138..2181
              /note="L2 repeat: matches 2648..2691 of consensus"
              /note="2282..2379
              /note="L2 repeat: matches 2582..2693 of consensus"
              /note="2483..2769
              /note="AluSc repeat: matches 1..287 of consensus"
              /note="2839..2962
              /note="MIR repeat: matches 20..170 of consensus"
              /note="3136..3434
              /note="AluSp repeat: matches 4..302 of consensus"
              /note="complement(3237..3650)
              /note="match: GSS: Em:AQ204224"
              /note="match: GSS: Em:AQ204224"
              /note="match: GSS: Em:AQ376380"
              /note="match: GSS: Em:AQ376380"
              /note="match: GSS: Em:AQ697506"
              /note="match: GSS: Em:AQ697506"
              /note="L2 repeat: matches 2575..2710 of consensus"

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:19:15 ; Search time 66.54 Seconds
(without alignments)
421.178 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 831
Sequence: 1 MAAGPSCGLVPFAFGKRLLLA.....ILKMWTDSEEFLEKLERI 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:**
1: sp_archaea:**
2: sp_bacteria:**
3: sp_fungi:**
4: sp_human:**
5: sp_invertebrate:**
6: sp_mammal:**
7: sp_mhc:**
8: sp_organelle:**
9: sp_phage:**
10: sp_plant:**
11: sp_rodent:**
12: sp_virus:**
13: sp_vertebrate:**
14: sp_unclassified:**
15: sp_virus:**
16: sp_bacteriaph:**
17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	822	98.9	162	4	O9GZM0
2	750	90.3	162	11	O9ERR7
3	744	89.5	162	11	O923V8
4	474	57.0	92	4	O9BS64
5	311	37.4	142	5	O9VVJ7
6	250.5	30.1	152	5	O9N4C6
7	226.5	27.3	158	10	O9SVL3
8	81	9.7	668	10	O22639
9	77	9.3	523	5	O9VQES
10	76.5	9.2	653	5	O94594
11	76.5	9.2	653	5	O25292
12	76.5	9.2	664	10	O39830
13	76	9.1	265	16	P73692
14	75.5	9.1	220	5	O94785
15	75.5	9.1	261	5	O03929
16	75.5	9.1	316	10	O9FHK0

17	75.5	9.1	653	5	O26936	O26936 trypanosoma
18	75.5	9.1	678	5	O26937	O26937 trypanosoma
19	75	9.0	456	12	O993Y0	O993Y0 spleena yel
20	74.5	9.0	665	10	O94IK4	O94IK4 scherffella
21	74.5	9.0	704	13	O73774	O73774 gallus gall
22	73	8.8	273	10	O9M2U5	O9M2U5 arabidopsis
23	73	8.8	274	10	O23248	O23248 arabidopsis
24	72.5	8.7	205	5	O95Y00	O95Y00 caenorhabd1
25	72.5	8.7	762	10	O9LE55	O9LE55 oryza sativ
26	72.5	8.7	830	5	O93176	O93176 caenorhabd1
27	72.5	8.7	838	5	O9U919	O9U919 caenorhabd1
28	72.5	8.7	838	5	O95QY9	O95QY9 caenorhabd1
29	72.5	8.7	1192	10	O9LYN8	O9LYN8 arabidopsis
30	72	8.7	383	3	O01678	O01678 pneumocysti
31	72	8.7	446	5	O60996	O60996 trypanosoma
32	72	8.7	620	5	O9NGM4	O9NGM4 drosophila
33	72	8.7	665	10	O9M4E8	O9M4E8 drosophila
34	72	8.7	777	10	O9M0X6	O9M0X6 cucumis sat
35	71.5	8.6	432	5	O9M6X2	O9M6X2 drosophila
36	71.5	8.6	514	5	O76958	O76958 leishmania
37	71.5	8.6	524	10	O91G99	O91G99 oryza sativ
38	71.5	8.6	663	5	O9GXR3	O9GXR3 leishmania
39	71.5	8.6	829	5	O9UB06	O9UB06 drosophila
40	71.5	8.6	872	5	O9NGQ0	O9NGQ0 drosophila
41	71.5	8.6	1175	5	O9YX8	O9YX8 drosophila
42	71.5	8.6	1175	5	O9VCX2	O9VCX2 drosophila
43	71.5	8.6	421	10	O92PR4	O92PR4 arabidopsis
44	71	8.5	560	3	O94667	O94667 schizosacch
45	71	8.5	560	3	O94667	O94667 schizosacch

ALIGNMENTS

RESULT	ID	Q9GZM0	PRELIMINARY:	PRT:	162 AA.
AC	O9GZM0	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)				
DT	01-MAR-2001 (TREMBLrel. 16, last annotation update)				
DE	DJ604K5.1 (15 KDA SELENOPROTEIN).				
GN	DJ604K5.1 OR SEP15.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]	SEQUENCE FROM N.A.			
RP	Pearce A.;				
RL	Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.				
RN	[2]	SEQUENCE FROM N.A.			
RP	Pubmed=10945981;				
RA	Kumaraswamy E., Malykh A., Korotkov R.V., Kozayavkin S., Hu Y.,				
RA	Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,				
RA	Hatfield D.L., Diamond A.M., Gladyshev V.N.;				
RT	"Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.				
RT	POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY.";				
DR	J. Biol. Chem. 275:35540-35547(2000).				
DR	EMBL; AL121989; CAC04186.1; -				
DR	EMBL; AF288991; AAG31556.1; -				
SO	SEQUENCE 162 AA; 17750 MW; 5894FBB423A089EE CRC64;				

Query Match 98.9%; Score 822; DB 4; Length 162;
Best Local Similarity 99.4%; Pred. No. 6,4e-79;
Matches 161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MAAGPSCGLVPFAFGKRLLATVLAQASAFGEFSSEACRELGSSNLLCSCDILGQFNL 60
DB 1 MAAGPSCGLVPFAFGKRLLATVLAQASAFGEFSSEACRELGSSNLLCSCDILGQFNL 60

QY 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162

RESULT 2
 ID Q9ERR7 PRELIMINARY; PRT: 162 AA.
 AC Q9ERR7:
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE 15 KDA SELENOPROTEIN.
 GN SEPI5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20519655; PubMed=10945981;
 RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozyavkin S., Hu Y.,
 RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
 RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
 RT "Structure-Expression Relationships of the 15-kDa Selenoprotein Gene.
 RT POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY."
 RT J. Biol. Chem. 275:35540-35547(2000).
 DR EMBL: AF288740; AAC31765.1; -.
 DR MGI:1927947; Sep15.
 SQ SEQUENCE 162 AA; 17738 MW; 6AB4B37D8C8B78447 CRC64;

Query Match 90.3%; Score 750; DB 11; Length 162;
 Best Local Similarity 92.0%; Pred. No. 2.5e-71;
 Matches 149; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162

RESULT 3
 ID Q923V8 PRELIMINARY; PRT: 162 AA.
 AC Q923V8:
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE 15 KDA CYTOSOLIC SELENOPROTEIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Roethlied D., Kyriakopoulos A., Behne D.;
 RT "A 15 kDa selenoprotein in several tissues of the rat."
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF90544; AAK73100.1; -.
 SQ SEQUENCE 162 AA; 17706 MW; 65E939968CB872AC CRC64;

Query Match 89.5%; Score 744; DB 11; Length 162;
 Best Local Similarity 92.6%; Pred. No. 1.1e-70;
 Matches 150; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 DB 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCGXKLGRRPVOAFAVRSRDKPKLFRGLQIKY 120
 QY 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162
 DB 121 VRGSDPVKLKLDNGNIAEELSIKWNMTDSVEEFLSEKLERI 162

RESULT 4
 ID Q9BS64 PRELIMINARY; PRT: 92 AA.
 AC Q9BS64:
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHEICAL 9.7 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RENAL ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BONE MARROW, CHRONIC MYELOGENOUS LEUKEMIA;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005294; AAH05294.1; -.
 DR EMBL: BC016359; AAH16359.1; -.
 SQ SEQUENCE 92 AA; 9669 MW; 59EABAD57FAD58 CRC64;

Query Match 57.0%; Score 474; DB 4; Length 92;
 Best Local Similarity 98.9%; Pred. No. 1.7e-42;
 Matches 91; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 DB 1 MAAGSGCLVPAFGKRLILATVLOAVSAFGAEFSSSEACRELGFSNNLCCSCDILGQFNL 60
 QY 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCG 92
 DB 61 LQIDPDCGCCGCEAOFETKLYAGAILLEVCG 92

RESULT 5
 ID Q9VVJ7 PRELIMINARY; PRT: 142 AA.
 AC Q9VVJ7:
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE CG7484 PROTEIN.
 GN CG7484.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champé M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butris K.C., Bussem D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajaji M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei V., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003523; AAF49314.1; -
DR FlyBase: FBgn0036745; CG7484.
SQ SEQUENCE 142 AA; 15965 MW; 01D06725BE3DB4DF CRC64;

Query Match 37.4%; Score 311; DB 5; Length 142;
Best Local Similarity 51.8%; Pred. No. 4.6e-25;
Matches 59; Conservative 17; Mismatches 38; Indels 0; Gaps 0;

QY 48 LCSSCDLLGQFNLLQJDPDCRCCEAEAFETKRLYAGAILVEYCGXKLGFRPOVQAFVRS 107
DB 1 MCSSCEKLDLDFDITRKPCQKCCCTLDQPPAQRFAKAILVEYCGXKLGFRPOVQAFVRS 60
QY 108 DKRFLEQLQIVYVGSPPVYKLLDNGNIAELSLTKNNTSVSEFLSEKLER 161
DB 61 GRPAFEPNLIQIVRGDLPVYKLLDASGRVQETLSITKNTDTVEEFETHLAK 114

RESULT 6
Q9N4C6 PRELIMINARY; PRT; 152 AA.
AC Q9N4C6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHEtical 17.6 KDA PROTEIN.
GN Y76B12C.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
CX NCBI_TaxID=6239;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cordes M.;
RT "The sequence of C. elegans cosmid Y76B12C.";
RT Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC024211; AAF36064.2; -
KW Hypothetical protein.
SQ SEQUENCE 152 AA; 17593 MW; 9CAA2F9C0ABE5B87 CRC64;

Query Match 30.1%; Score 250.5; DB 5; Length 152;
Best Local Similarity 40.7%; Pred. No. 1.2e-18;
Matches 61; Conservative 26; Mismatches 54; Indels 9; Gaps 5;

QY 17 LLIATLQAVSAGFA----EFSSEACRELGFS-SNLCSSCDLLGQFNLLQJDPDCRCG 71
DB 6 LLIAAASVSP--FGVEYERKIDYECCKAGFENETKCGICERLSYHLETLTDCGCC 63
QY 72 OEBAQEFKRLYAGAILVEYCGXKLGFRPOVQAFVRSRDKPLFNG-LQIKVRSPPVRL 130
DB 64 IKREERKHEK-YPTALEYCECLARFPQVAVHDMARFGKTKVKKHVRQVVAL 122
QY 131 LDDNGNIAELSLTKNNTSVSEFLSEKLE 160
DB 123 KDAEFKTEKLVSEKWDPTLLIDFFNQMLE 152

RESULT 7
Q9STL3 PRELIMINARY; PRT; 158 AA.
AC Q9STL3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F3F20.17 PROTEIN.
GN F3F20.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Ruiz L., Kowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Lueros S., Schwartz J., Shinn P., Toriumi M., Vysotskaya V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007153; AAD30621.1; -
DR InterPro: IPR000345; Cytochrome_bbind.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOMN.1.
SQ SEQUENCE 158 AA; 17901 MW; A181CB37DA4A33B8B CRC64;

Query Match 27.3%; Score 226.5; DB 10; Length 158;
Best Local Similarity 32.9%; Pred. No. 4.3e-16;

```

	Query Match	9.7%	Score 81;	DB 10;	Length 668;	
	Best Local Similarity	22.5%;	Pred. No. 5.5;			
	Matches	40;	Conservative	26;	Mismatches 62;	Indels 50; Gaps 7.
OY	5	PSCGLVPFAGKRLLIATVQ-----ANSACAEESSEACRELGRSSNLLCSCDILLOGF	58			
		: : : : : : : : : : : : : : : : : :				
Db	446	FRNTVIPF--KKSOVFYYODOOTTSIOVEFGERS-----LTRDCRLLGKF	490			
OY	59	NLLQDPDPCRGCCOEAPFETKRLAYALILECGAKLGFPOVAQFAVRSDPKLFELRGLQI	118			
		: : : : : : : : : : : : : : : : : : : :				
Db	491	DLSCGPAPRCGTALETFEEVD---ANGLINVKAKDKDGKGSEKITITNEKGRLL	541			
OY	119	KYVRGSDDVLKLIDONGINAELSILKNUTD---SVEEP-----LSKELE	160			
		: : : : : : : : : : : : : : : : : :				
Db	542	----SQEELIERAVREBKOPAFEBEKRYKERIDARNSLETIVYMKNQOVSDPKDLADKLE	595			
RESULT	9					

09VQE9

ID	09VQE9	PRELIMINARY:	PRT:	523 AA.
AC	09VQE9:			
DT	01-MAY-2000 (TREMBLrel. 13)	(Created)		
DT	01-MAY-2000 (TREMBLrel. 13,	Last sequence update)		
DE	01-JUN-2001 (TREMBLrel. 17,	Last annotation update)		
DZ	CG11542 PROTEIN.			
GN	CG11542.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephydroidea; Drosophilidae; Drosophila.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY:			
RA	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celisner S.E., Holt R.W., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galie R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers V.-H.C., Blazer J.C., Champe M., Pfeiffer B.D.,			
RA	Ban K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu P.V., Bertram B.P., Bhattacharya D., Bolshakov S.,			
RA	Beeson K.Y., Benson A., Bonin J., Bouck J., Brockstein P., Brotler P.,			
RA	Borkova D., Botchan M.R., Bouton J., Brookstein P., Brothier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Douc L.E., Downes N., Dugan-Kocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferrara C., Fertiera S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,			
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwan C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison K.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosherrefl A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puril V., Reese M.G.,			
RA	Reinhart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shen B.C., Sliden-Klamas I., Simpson M., Skupski M.P., Smith T.,			
RA	Splter E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskys R., Teicor C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,			
RA	Xe J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.);			
RT	"The genome sequence of Drosophila melanogaster."			
RL	Science 287:2185-2195(2000).			
DR	EMBL: AE003582; AAF51225.1; -			
DR	FLYBase: FBgn0031447; CG11542.			
DR	InterPro: IPRO01440; TPR.			
DR	Pfam: PF00515; TPR; 1.			
SO	SEQUENCE 523 AA; 57320 MW; BABA560198C3E6DA CRC64;			

Query Match 9.3%; Score 77; DB 5; Length 523;
Best Local Similarity 27.0%; Pred. NO. 11;
Matches 27; Conservative 17; Mismatches 42; Indels 14; Gaps 5;

DB **18** LTAATGAAVSAGFAEGSSBACRELFFSSLTSCSDDILGOFNLTOLDPDCRGCCOEAOE 77
 |:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-
DB 151 ILITLMALMAL-----ICKETAIFATLLLCGLGDVLSPVGRENSDKVCDSISIGLSLF 203
 ::::|||||||:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-

OY 78 ETKKIYAG-AILE---VGCX--KLGRRFOQV-AAPRSKP 110
 :::::|||:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-|:-

Db 204 NFQRFRSLISTGIETLLCGLYCRISLPDPPTATAFSAADNP 243

```

RESULT 10
O94594 PRELIMINARY: PRT: 653 AA.
AC O94594;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEAT SHOCK PROTEIN 70.
GN HSP70.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEM75;
RX MEDLINE=97172530; PubMed=9020174;
RA Oujada L., Soto M., Alonso C., Requena J.M.;
RT *Analysis of post-transcriptional regulation operating on
RT transcription products of the tandemly linked Leishmania infantum
RT hsp70 genes.*;
RL J. Biol. Chem. 272:4493-4499(1997).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL: Y08020; CA69282.1; -.
DR HSSP: Y08019; CA69281.1; -.
DR HSSP: P08107; IH00.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR ATP-binding; Heat shock.
KW SEQUENCE 653 AA; 71078 MW; 908669AB8B20BBF CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 5; Length 653;
Best Local Similarity 27.9%; Pred. No. 16;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDLLGQFNLDLPDPCGCCGEAEQFETKLYAGAILLEVCCKLGRFPQVAFVRS 107
DB 451 MKKDCHLGTFDLGSGIPRPRGRLPQIEVTFD---LDANGILNVAEKEGTGRKNQITTN 507
OY 108 DKPKLFRGLQIKYVRGSDPYLKLDD 133
DB 508 DKGRL-----SKDEIERMVND 523

RESULT 11
O25292 PRELIMINARY: PRT: 653 AA.
AC O25292;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HEAT-SHOCK PROTEIN (FRAGMENT).
GN HSP70.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEM75 /ZYMODEME 1;
RX MEDLINE=96352058; PubMed=8728991;
RA Oujada L., Requena J.M., Soto M., Alonso C.;
RT *during canine viscero-cutaneous leishmaniasis the anti-Hsp70
RT antibodies are specifically elicited by the parasite protein.*;
RL Parasitology 112:277-284(1996).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL: X85798; CA59793.1; -.
DR HSSP: P08107; IH00.

```

```

DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR ATP-binding.
KW NON_TER 653
FT SEQUENCE 653 AA; 71163 MW; A4EAB21C07F5E08E CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 5; Length 653;
Best Local Similarity 27.9%; Pred. No. 16;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDLLGQFNLDLPDPCGCCGEAEQFETKLYAGAILLEVCCKLGRFPQVAFVRS 107
DB 451 MKKDCHLGTFDLGSGIPRPRGRLPQIEVTFD---LDANGILNVAEKEGTGRKNQITTN 507
OY 108 DKPKLFRGLQIKYVRGSDPYLKLDD 133
DB 508 DKGRL-----SKDEIERMVND 523

RESULT 12
O39830 PRELIMINARY: PRT: 664 AA.
AC O39830;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BIP ISOFORM A.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CENTURY 84; TISSUE-LEAF;
RX MEDLINE=95218610; PubMed=7766051;
RA Kalinski A., Rowley D.L., Loer D.S., Foley C., Buta G., Herman E.M.;
RT *Binding-protein expression is subject to temporal, developmental and
RT stress-induced regulation in terminally differentiated soybean
RT organs.*;
RL Planta 195:611-621(1995).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR EMBL: U08384; AA81956.1; -.
DR HSSP: P19120; 3HSC.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR001023; HSP70.
DR InterPro: IPR000671; Hydrgrn_uptake.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR ATP-binding.
KW SEQUENCE 664 AA; 73336 MW; E1E2ABCB2086E3D6 CRC64;
SQ

Query Match 9.2%; Score 76.5; DB 10; Length 664;
Best Local Similarity 22.2%; Pred. No. 16;
Matches 45; Conservative 26; Mismatches 69; Indels 63; Gaps 8;

OY 5 PSCGIVAFGRKLLATVQ-----AVSAFGAFSSSEACRELGFSNLLCSCDILGQF 58
DB 444 PRVTVPIT--KRSQVFTTYQDOQSTVSIOVFEEGRS-----LTKDCRLGRF 488
OY 59 NLLDLPDPCGCCGEAEQFETKLYAGAILLEVCCKLGRFPQVAFVRS DKPKLFRGLQI 118
DB 489 ELGIPRPRGRLPQIEVTFD---ANGILNVAEDKGTGKSEKITITNEKRLSOEEIE 545

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 16:30:50 ; Search time 24.1 Seconds

(without alignments)
260.272 Million cell updates/sec

Title: US-09-676-718A-1

Perfect score: 831
Sequence: 1 MAAGPSGCLVPAGFKRLLLA.....ILKWNDSVEEFISEKLERI 162

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	821	98.8	162	SE15_HUMAN	060613 homo sapien
2	756	91.0	162	SE15_RAT	0922v8 rattus norv
3	749	90.1	162	SE15_MOUSE	09e2r7 mus musculu
4	82	9.9	485	STE_RHIME	P15189 rhizobium m
5	81	9.7	653	HS70_LEIDO	P17804 leishmania
6	81	9.7	680	HS70_TRYCR	P05456 trypanosoma
7	80.5	9.7	661	HS74_TRYBB	P11445 trypanosoma
8	80	9.6	269	THRG_CAUCR	09e746 callobacter
9	80	9.6	652	HS70_LE1AM	007437 leishmania
10	80	9.6	983	PTPN_RAT	063359 rattus norv
11	78	9.4	290	BIP1_TOBAC	003681 nicotiana t
12	78	9.4	292	BIP2_TOBAC	003682 nicotiana t
13	76.5	9.2	557	ILVD_BACSU	P51785 bacillus su
14	76	9.1	668	BIP5_TOBAC	003685 nicotiana t
15	74.5	9.0	516	HS70_TOBAC	P14834 leishmania
16	74.5	9.0	667	GR78_APLCA	P16936 aplysia cal
17	74.5	9.0	670	GR78_YARLI	099170 yarrowia li
18	74.5	9.0	684	FBL1_CHICK	073775 gallus gall
19	74	8.9	264	THRG_XYFLA	09p695 xyella fas
20	74	8.9	663	GR78_SCHPO	P36604 schizosacch
21	74	8.9	979	PTPN_MOUSE	060673 mus musculu
22	73	8.8	293	BIP8_TOBAC	003686 nicotiana t
23	73	8.8	1120	MDP_RICPR	005955 rickettsia
24	72.5	8.7	228	HS70_LE1BR	P27894 leishmania
25	72.5	8.7	667	BIP4_TOBAC	003684 nicotiana t
26	72.5	8.7	830	DYN1_CAEEL	P39055 caenorhabdl
27	72	8.7	663	BIP2_MAIZE	P24067 zea mays (m
28	71.5	8.6	643	HS71_DROME	P02825 drosophila
29	71.5	8.6	656	HS7C_DROME	P29844 drosophila
30	71.5	8.6	668	BIP_SPIOI	Q42434 spinacia ol
31	71	8.5	890	KPCN_HUMAN	Q94806 homo sapien
32	71	8.5	979	PTPN_BOVIN	P56722 bos taurus
33	71	8.5	979	PTPN_HUMAN	Q16649 homo sapien

34	70.5	8.5	212	1	DSBA_BUCAT	P57505 buchnera ap
35	70.5	8.5	432	1	SYN_LACDE	P54262 lactobacill
36	70.5	8.5	643	1	HS71_SCHPO	Q10265 schizosacch
37	70.5	8.5	659	1	YH5_YEAST	P38760 saccharomyc
38	69.5	8.4	640	1	HS74_ANOAL	P41827 anopheles a
39	69	8.3	265	1	THRG_PSEAE	Q916b4 pseudomonas
40	69	8.3	666	1	BIP_LYCES	P49118 lycopersico
41	69	8.3	668	1	BIP2_ARATH	Q39043 arabidopsis
42	69	8.3	669	1	BIP1_ARATH	Q91kr3 arabidopsis
43	69	8.3	677	1	SPOT_RHEIN	P43811 haemophilus
44	69	8.3	682	1	GR78_YEAST	P16474 saccharomyc
45	69	8.3	715	1	LCCL_LACLA	Q9cjb8 lactococcus

ALIGNMENTS

RESULT 1
ID SE15_HUMAN STANDARD: PRT; 162 AA.
AC 060613; Q9NR01; Q9BS64; Q9GZM0;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kDa selenoprotein precursor.
GN SEP15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 98-106; 123-127 AND 146-158.
RX MEDLINE-98204881; PubMed-9535873;
RA Gladyshev V.N., Jeang K.-T., Mooton J.C., Hatfield D.L.,
RT "A new human selenium-containing protein. Purification,
RL characterization, and cDNA sequence."
RN J. Biol. Chem. 273:8910-8915(1998).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE-20519655; PubMed-10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozayvkin S., Hu Y.,
RA Kwon S.Y., Moustafa M.E., Carlson B.A., Berry M.J., Lee B.J.,
RT Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RL "Structure-expression relationships of the 15-kDa selenoprotein gene.
RN Possible role of the protein in cancer etiology."
RN J. Biol. Chem. 275:35540-35547(2000).
RN [3]
RN SEQUENCE FROM N.A.
RA Ryu M., Moon E.;
RT "The human 15-kDa selenoprotein gene: characterisation of the genomic
RN structure and functional analysis of the promoter."
RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RA Pearce A.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC TISSUE=Bone marrow, and Kidney;
RP Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: SEEMS TO BE EITHER OLIGOMERIC OR HETEROOIGOMERIC.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- MASS SPECTROMETRY: MW=14870; METHOD-Electrospray; RANGE=7-152.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AF051994; AAC15478.1; -
DR EMBL: AF288991; AAG31556.1; -
DR EMBL: AF288992; AAG31557.1; -
DR EMBL: AF267986; AAF78966.1; ALT_SEQ.
DR EMBL: AF267982; AAF78966.1; JOINED.
DR EMBL: AF267983; AAF78966.1; JOINED.
DR EMBL: AF267984; AAF78966.1; JOINED.
DR EMBL: AF267985; AAF78966.1; JOINED.
DR EMBL: AL121989; CAC04186.1; ALT_SEQ.
DR EMBL: BC005294; AAH05294.1; ALT_TERM.
DR EMBL: BC016359; AAH16359.1; ALT_TERM.
DR MIM: 606254; -
KW Selenium; selenocysteine; signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17743 MW; 463EACBA23B4CDBE CRC64;

Query Match 98.88; Score 821; DB 1; Length 162;
Best Local Similarity 98.88; Pred. No. 3.9e-75;
Matches 160; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
QY 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
QY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 2
SE15_RAT STANDARD; PRT; 162 AA.
AC Q923V8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kda selenoprotein precursor.
CN SEP15.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Roethlislein D., Kyriakopoulos A., Behne D.;
RT "A 15 kda-selenoprotein in several tissues of the rat."
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF390544; AAK73100.1; ALT_SEQ.
KW Selenium; selenocysteine; signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17775 MW; 73985E968C9F36FC CRC64;

Query Match 91.08; Score 756; DB 1; Length 162;
Best Local Similarity 92.68; Pred. No. 1.2e-68;

Matches 150; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
QY 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
QY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 3
SE15_MOUSE STANDARD; PRT; 162 AA.
AC Q9ERR7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 15 kda selenoprotein precursor.
CN SEP15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-20519655; Pubmed-10945981;
RA Kumaraswamy E., Malykh A., Korotkov K.V., Kozyavkin S., Hu Y.,
RA Kwon S.Y., Mustafa M.E., Carlson B.A., Berry M.J., Lee B.D.,
RA Hatfield D.L., Diamond A.M., Gladyshev V.N.;
RT "Structure-expression relationships of the 15-kda selenoprotein gene.
RT Possible role of the protein in cancer etiology."
RL J. Biol. Chem. 275:35540-35547(2000).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF288740; AAG31765.1; ALT_SEQ.
DR MGD: MGI:1927947; Sep15.
KW Selenium; selenocysteine; signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 162 15 KDA SELENOPROTEIN.
FT SE_CYS 93 93
SQ SEQUENCE 162 AA; 17731 MW; 771E047D8C9F3C17 CRC64;

Query Match 90.18; Score 749; DB 1; Length 162;
Best Local Similarity 91.48; Pred. No. 6.2e-66;
Matches 148; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
DB 1 MAAGSGCLVAFGKRLATVLAQVSAFGAEFFSEACRELGFSNNLTCSSCDLLGQFNL 60
QY 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
DB 61 LQLPDPCRGCCOEAEQFETKRLIYAGAILLEVCGXKLGREPVOQAFVRSKPKLFRGLQIKY 120
QY 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162
DB 121 VRGSDPVKLDDNGNIAEELSIKWNDSVEEFLSEKLERI 162

RESULT 4


```

DT      01-FEB-1991 (Rel. 17, Last annotation update)
DE      Heat shock 70 kDa protein.
GN      HSP70.
OS      Trypanosoma cruzi..
OC      Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX      NCBI_TaxID=5693;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=86157703; PubMed=2831499;
RA      Requena J.M., Lopez M.C., Jimenez-Ruyiz A., de la Torre J.C.,
RT      Alonso C.;
RL      "A head-to-tail tandem organization of hsp70 genes in Trypanosoma
        cruzi."
CC      Nucleic Acids Res. 16:1393-1406(1988).
CC      -I SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X07083; CA830115.1; -.
DR      PIR; A28531; A28531.
DR      PIR; S06158; S06158.
DR      HSSP; P08107; 1HDO.
DR      InterPro; IPR001023; HSP70.
DR      Pfam; PF00012; HSP70_1.
DR      PRINTS; PR00301; HEATSHOCK70.
DR      PROSITE; PS00297; HSP70_1; 1.
DR      PROSITE; PS00329; HSP70_2; 1.
DR      PROSITE; PS01036; HSP70_3; 1.
DR      ATP-binding; Heat shock; Multigene family.
SQ      SEQUENCE 680 AA; 73809 MW; D102000F5B7A8D82 CRC64;

Query Match          9.7%; Score 81; DB 1; Length 680;
Best Local Similarity 27.8%; Freq. No. 2;
Matches 37; Conservative 16; Mismatches 52; Indels 28; Gaps 4;

OY      48 ICSSCDLIGQFMILDLDPDCRCGCEAEAFERKRLTYAGALIEVCGKKIGRFPQVOAFVRS 107
           | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB      452 MTKDCHILGTFLSGLSIPPPRGVPDIETFD---LDANGILNVSAEKGKTGRNDIVLTN 508
           | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY      108 DKPKTFRGLQIQTVR-----GSDPYVLKLDDNG-----NIAEL--S 142
           | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB      509 DKGRLSRAEIEMVHEAKYEADDDVDVQIDAKNGLENTAFSRMNAVNPVAVGKIEEA 568
           | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
OY      143 ILKNWTDSEVEFL 155
           | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB      569 DKTTTSAYEAL 581

RESULT 7
HS74_TRYBB ID HS74_TRYBB STANDARD: PRT: 661 AA.
AC P11145;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 01-MAY-1992 (Rel. 22, Last annotation update)
DE Heat shock 70 kDa protein 4 (HSP70).
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=87089827; PubMed=3796613;
RA Glass D.J., Polvere R.I., van der Ploeg L.H.T.;
RT "Conserved sequences and transcription of the hsp70 gene family in
   Trypanosoma brucei".
RT Mol. Cell. Biol. 6:4657-4666(1986).
```

```
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M16997; AAA30204.1; -.
DR PIR; A25398; A25398.
DR HSSP; P19120; IHPM.
DR InterPro; IPR001023; HSP70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR App-binding; Heat shock; Multigene family.
KW SEQUENCE 661 AA; 71433 MW; F9F927003E595A13 CRC64;
SQ
Query Match          9.7%; Score 80.5; DB 1; Length 661;
Best Local Similarity 24.7%; Pred. No. 2.2;
Matches 36; Conservative 16; Mismatches 55; Indels 39; Gaps 4;
QY 48 LCSSCDLLGQNPLQLDPCRCGCCCEEAQFTKKRLVAGAILVECGXKRGPPVOVAFYRS 107
   : |||||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 452 MTNKHCHLGTFDLGSIPAPRGVPQIEVTPD---LDANGILLSVSAAEKGGTKRNDIYVTN 508
QY 108 DKPKLFH-----GLQ-----IKYVGSDPYVLKLDDN 134
   ||::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 509 DKGRLSKADIERMVSDAKYEAEDEKAHYHXIDAKNGLENTAFVSMKMTINDPNVAGKLDDA 568
OY 135 GNIAEELSI---LKWNTDSVEFLSE 157
Db 569 DKNAVTTAVEERALRWLNNDQASLEE 594
-----
RESULT      8
THIG_CAUCR  THIG CAUCR      STANDARD;      PRT;      269 AA.
AC 09A746;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Thiazole biosynthesis protein thig.
GN THIG OR CC1880.
OC Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
CC Caulobacter.
CN NCBI_TaxID=69394;
CX [1]
SEQUENCE FROM N.A.
PP STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11295647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: Required for the synthesis of the thiazole moiety of
CC thiamine (By similarity).
CC -1- COFACTOR: FMN (Potential).
CC -1- PATHWAY: Thiamine biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE THIG FAMILY.
-----
```



```

DB 71 PRNTVPT--KKSQVFTTYDDOQTTSIOVFEGERS-----LTKDCRLGKF 115
QY 59 NLLQDLPDRCGCCOEAEQFETKLYAGALILEGCKXKRGFPQOAVRSPKPLFRGLQI 118
DB 116 DLGTIAPAPRGPTQIEVTEFVD--ANGILINVAEDKGKSGSKRITTTNDKGL-----166
QY 119 KYVKGSDPVLKLLDNGNINAEELSLKMNND---SVEEF-----LSEKLE 160
DB 167 ----SQEIERMVRREAEFEAEEDKKYKERIDANSLSLETYYNNKNOINDKDLADKLE 220

```

RESULT 12

```

BIP2_TOBAC STANDARD; PRT; 292 AA.
ID BIP2_TOBAC STANDARD; PRT; 292 AA.
AC Q03682;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Luminal binding protein 2 (BIP 2) (78 kDa glucose-regulated protein
DE homolog 2) (GRP 78-2) (Fragment).
GN BIP2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN NM (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;
RA Denecke J., Goldman M.H., Demolder J., Seurlinck J., Botterman J.;
RT "The tobacco luminal binding protein is encoded by a multigene
RT family."
RL Plant Cell 3:1025-1035(1991).
RN RM (2)
RP ERRATUM.
RA Denecke J., Goldman M.H., Demolder J., Seurlinck J., Botterman J.;
RL Plant Cell 3:1251-1251(1991).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X60059; CAA42661.1; .
DR PIR: S21878; S21878.
DR PIR: P00262; P00262.
DR HSSP: P19120; INCG.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70.1.
DR PROSITE: PS00297; HSP70.1; PARTIAL.
DR PROSITE: PS00329; HSP70.2; PARTIAL.
DR PROSITE: PS01036; HSP70.3; PARTIAL.
DR PROSITE: PS00014; ER_TARGET; 1.
DR ATP-binding; Endoplasmic reticulum; Multigene family.
FT NON_TER 1 1
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 289 292 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 292 AA: 32260 MW: 667A4F5107C6E2D5 CRC64;

```

Query Match 9.4%; Score 78; DB 1; Length 292;
 Best Local Similarity 23.0%; Pred. 1.5;
 Matches 41; Conservative 24; Mismatches 63; Indels 50; Gaps 7;

```

QY 5 PSCGLVPARGKRLLLATVLD-----AVSAFGAEFSSSEACRELGFSNNLICSSCDLLGOF 58
DB 71 PRNTVPT--KKSQVFTTYDDOQTTSIOVFEGERS-----LTKDCRLGKF 115
QY 59 NLLQDLPDRCGCCOEAEQFETKLYAGALILEGCKXKRGFPQOAVRSPKPLFRGLQI 118
DB 116 DLGTIAPAPRGPTQIEVTEFVD--ANGILINVAEDKGKSGSKRITTTNDKGL-----166
QY 119 KYVKGSDPVLKLLDNGNINAEELSLKMNND---SVEEF-----LSEKLE 160
DB 167 ----SQEIERMVRREAEFEAEEDKKYKERIDANSLSLETYYNNKNOINDKDLADKLE 220

```

RESULT 13

```

ILVD_BACSU STANDARD; PRT; 557 AA.
ID ILVD_BACSU STANDARD; PRT; 557 AA.
AC P51785;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD) (vegetative protein
DE 110) (VEG110).
GN ILVD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN NM (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
RA Serrro P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the sera and kdg loci cloned in a yeast artificial chromosome."
RL Microbiology 142:2005-2016(1996).
RN RM (2)
RP SEQUENCE OF 1-15.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis."
RL Electrophoresis 18:1451-1463(1997).
CC -1- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate -> 3-methyl-2-
CC oxobutanoate + H(2)O.
CC -1- COFACTOR: BINDS 1 4FE-4S CLUSTER (POTENTIAL).
CC -1- PATHWAY: FOURTH STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L77246; AAA96629.1; .
DR EMBL: Z99115; CAB14105.1; .
DR Subtilist: BG01532; ILVD.
DR InterPro: IPR000581; ILVD_EDD.
DR Pfam: PF00920; ILVD_EDD.1.
DR Pfam: PD002691; ILVD_EDD.1.
DR PROSITE: PS00886; ILVD_EDD_1; 1.
DR PROSITE: PS00887; ILVD_EDD_2; 1.
DR Branched-chain amino acid biosynthesis; Lyase; Iron-sulfur; 4Fe-4S;
KW Complete proteome.
FT INIT_MER 0 0
FT METAL 121 121 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 194 194 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 557 AA: 59417 MW: CAC59717EB9D7EDA CRC64;

```

Query Match 9.28; Score 76.5; DB 1; Length 557;
Best Local Similarity 21.5%; Pred. No. 4.5;
Matches 37; Conservative 33; Mismatches 75; Indels 27; Gaps 7;

OY 1 MAAGPSCVPAFCRLLATVLOAVSAPGA-ESSSEACRELGSNNLSCDCLLOFN 59
DB 142 VSGGPMAGRTSGRTKISLSVFEVGAYOGAKINENELDEQFGCPTGSGS--GMFT 199
OY 60 LLADPDCRCGCOEAFETKLYAGAILLEVCGXKLGFRPQVAFYRSKPKLFRGLQIK 119
DB 200 ANSMN-----CLSEA-----LGLALPONGTILATSPERKEFYKSAQAQIMETIR-K 244
OY 120 YVSGSDPV-LKLDL-----NGNTAEISLTKMNTDVEEFLSEKLERI 162
DB 245 DIKPRDVTYKADINAFALDMALGSGTNTVLTALANAGVEYSLERINEV 296

RESULT 14

BIP5_TOBAC STANDARD; PRT; 668 AA.

ID BIP5_TOBAC STANDARD; PRT; 668 AA.
AC Q03685;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Luminal binding protein 5 precursor (Bip 5) (78 kDa glucose-regulated protein homolog 5) (GRP 78-5).
GN BIP5.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eunasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;
RA Denecke J., Goldman M.H., Demolder J., Seurinck J., Botteman J.;
RT "The tobacco luminal binding protein is encoded by a multigene family."
RL Plant Cell 3:1025-1035(1991).
RN [2]
RP ERRATUM.
RA Denecke J., Goldman M.H., Demolder J., Seurinck J., Botteman J.;
RL Plant Cell 3:1251-1253(1991).
CC -!- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60058; CAA42660.1; -
DR PIR: S21880; S21880.
DR PIR: J01361; J01361.
DR HSSP: P19120; 3HSC.
DR InterPro: IPR000866; ER_target.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
KM APP-binding; Heat shock; Multigene family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 668 LUMINAL BINDING PROTEIN 5.

FT CARBOHYD 617 617 N-LINKED (GLCNAC. ...) (POTENTIAL).
FT SITE 665 668 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 668 AA; 73744 MW; 7980231A91DC590 CRC64;

Query Match 9.18; Score 76; DB 1; Length 668;
Best Local Similarity 23.08; Pred. No. 6.2;
Matches 41; Conservative 23; Mismatches 64; Indels 50; Gaps 7;

OY 5 PSCGLVPAFCRLLATVLO-----AVSAFGAFSSSEACRELGSNNLSCDCLLOGF 58
DB 447 PRNVIFP--KSGVFTTYDQQTIVTISVFEGRS-----LTKCRLLGKF 491
OY 59 NLADPDCRCGCOEAFETKLYAGAILLEVCGXKLGFRPQVAFYRSKPKLFRGLQI 118
DB 492 DLGIAAPAPRGTPQIEVTFEVD--ANGILNVKAEDKASGSEKITITNDKGRF----- 542
OY 119 KYVSGSDPV.LKLDNDNINIAEISLTKMNTD--SVEEF-----LSEKLE 160
DB 543 ----SDEIERMVAEAEFAEDKKKVERIDARNSLETYYNMRNQINDXDKLADKLE 596

RESULT 15

HS70_LEIMA STANDARD; PRT; 516 AA.

ID HS70_LEIMA STANDARD; PRT; 516 AA.
AC P14834;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Heat shock 70 kDa protein (Fragment).
GN HSP70.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WR300;
RX MEDLINE=89041551; PubMed=3186441;
RA Lee M.G., Atkinson B.L., Giannini S.H., van der Ploeg L.H.T.;
RT "Structure and expression of the hsp 70 gene family of Leishmania major."
RL Nucleic Acids Res. 16:9567-9585(1988).
RN [2]
RP Nucleic Acids Res. 16:9567-9585(1988).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36675; AAA29251.1; -
DR PIR: S06443; S06443.
DR HSSP: P08109; 1CKR.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
KM APP-binding; Heat shock; Multigene family.
FT NON_TER 516
FT SEQUENCE 516 AA; 56535 MW; 36F83ACE952452 CRC64;

Query Match 9.08; Score 74.5; DB 1; Length 516;
Best Local Similarity 33.88; Pred. No. 6.5;
Matches 22; Conservative 9; Mismatches 31; Indels 3; Gaps 1;

OY 48 LCSSCDLGGFNILQDPDCRCGCOEAFETKLYAGAILLEVCGXKLGFRPQVAFYRS 107
DB 452 MTRKCHLGTGTDLSGIRPARGVQIEVTD--LDANGILNVSAEKGCKGRNQITITN 508

OY 108 DKPKL 112
11:1
Db 509 DKGRL 513

Search completed: September 18, 2002, 16:36:05
Job time: 315 sec

THIS PAGE BLANK (USPTO)

[illegible]

```

RESULT      2
US-08-493-754A-5
; Sequence 5, Application US/08493754A
; Patent No. 582138
; GENERAL INFORMATION:
; APPLICANT: Godwin, Andrew K.
; TITLE OF INVENTION: No. 582133bel Gene Associated with Suppression
; TITLE OF INVENTION: of Tumor Development
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/493,754A
; FILING DATE: 22-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hagan, Patrick J.
; REGISTRATION NUMBER: 27,643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-493-754A-5

Query Match          9.6%; Score 79.5; DB 2; Length 661;
Best Local Similarity 20.3%; Pred. No. 0.35;
Matches 59; Conservative % 28; Mismatches 64; Indels 139; Gaps

QY      8 CLVP-----ARGKLLATVYLQAV-----SANGAEFS-----EACRELGFS 44
       |||||         |||||:: ::| : | : | : | : | : | : | : | : | : | : |
Db      89 CLVPIONTDGIAMLYVEGKRLLVVSVTQFIPLSLQTLRTLTENKKDDSSIRIDIPCKPLSPG 148
       | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      45 SNLSCSDL-----LQPFN-----LQDPDRCCQG-----              73
       | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      149 EVLGGTSRLDASRYDAIVYLGDRFHLESIMINPELEAFQYDYSRKLTRFYEDHLM 208

```

```

QY 74 ----- -AGSET---KRLYAGALIEVCXKL 95
Db 209 KKNRIGSTIARCKCTTGGTLGIGTGROGNLKVVEELPAOLERRKKRLRVLLSEIPEKL 268
QY 96 GRPPOVOAFVRSDPKL-----FRGLIKTVRSGDPV 127
Db 269 AMFEVDICWVAVACPRLSIDMGTFPPKPLLYPEFLAVALDNYSKFFCLOI---TGQWTI 325
QY 128 LKLL----- -DNGNIAEELSLIKKNTDSVEEFLSEKLERI 162
Db 326 IRMLPFWLGRILIMKRYVRNGNGMLI-LLSKPRHSHSELSTFNEDEAKKRI 374

```

```

RESULT      3
US-08-441-139-7
: Sequence 7, Application US/08441139
: Patent No. 5773245
:
: GENERAL INFORMATION:
: APPLICANT: Wiltup, Dr. Karl D.
: APPLICANT: Robinson, Anne S.
: TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
: TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SULLY, SCOTT, MURPHY & PRESSER
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: NY
: COUNTRY: USA
: ZIP: 11530
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/441,139
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/089,997
: FILING DATE: 06-JUL-1993
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Digiglio, Frank S.
: REGISTRATION NUMBER: 31,346
: REFERENCE/DOCKET NUMBER: 8646
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 516-742-4343
: TELEFAX: 516-742-4356
:
: TELEX: 230 901 SANS UR
:
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 663 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
US-08-441-139-7

```

```

Query Match          8.9%; Score 74; DB 1; Length 663;
Best Local Similarity 27.7%; Pred. No. 1.7;
Matches    26; Conservative   17; Mismatches   37; Indels   14; Gaps    3;

OY      53 DILGQFNLLDPPDCRGCCQEAEQFETKYLVAIGALLEVCG-KLGRFPVOQAFAVSDDRKR 111
         |||::||:| | | | | | | | | | | : | | | : | | | : | | | :
Db       401 NLLGFEDLRGPAPRPAQPQILEVEEV--ANGVLTSAYVDKSGRKREKLVIKNDRGR 537
         |||::||:| | | | | | | | | | | : | | | : | | | : | | | :

OY      112 LFRGLIKYVRGSDPVLKLLDNNGNIAEELSLTK 145
         | | | | | | | | | | | | | | | | | | | |
Db       538 L-----SEEDIERNVKAEFEFAEDBKTLTK 561
         | | | | | | | | | | | | | | | | | | | |

RESULT      4
```


;; EARLIER FILING DATE: July 21, 1998
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: Microsoft Office 97
;; SEQ ID NO 31
;; LENGTH: 1072
;; TYPE: PRF
;; ORGANISM: Saccharomyces cerevisiae
US-09-357-231-31

Query Match 7.8%; Score 65; DB 4; Length 1072;
Best Local Similarity 26.8%; Pred. No. 43;
Matches 37; Conservative 27; Mismatches 52; Indels 22; Gaps 8;

OY 22 VLAQVAFGAEFSSACRELGFS--NLICSSCDLLGQFNLLDLPDRCGCCQEEA-QPE 78
Db 322 IVHNPARGEE-DNACCLKNGVISDSVLPNAILDGLRFT--KDPVDEFGVYVKADKII 378
OY 79 TKKLYAGALIEVCGKKGKLFPOVQAFV-RSDPKLFRGLQIYVKGSDPYLKLDDNGNI 137
Db 379 IKYLNNTGNLLASQIRHSYP---FCWRSDPPLLYRSVPAMFVRYKNIYPMOLD----- 429
OY 138 AEELSLK--NNTDSVEE 153
Db 430 ----SVKSHWVPNTIKE 443

RESULT 10
US-09-442-631-2
; Sequence 2, Application US/09442631
; Patent No. 6300489
; GENERAL INFORMATION:
; APPLICANT: OH, BOONG-JUN
; APPLICANT: KO, MOON KYUNG
; APPLICANT: SHIN, BYONGCHUL
; APPLICANT: CHUNG, CHANG HO
; TITLE OF INVENTION: SMALL AND CYSTEINE RICH ANTITUMORAL DEFENSIN AND
; TITLE OF INVENTION: THIONIN-LIKE PROTEIN GENES HIGHLY EXPRESSED IN THE
; FILE REFERENCE: 1942/44
; CURRENT APPLICATION NUMBER: US/09/442,631
; CURRENT FILING DATE: 1999-11-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 83
; TYPE: PRF
; ORGANISM: Capsicum annuum
US-09-442-631-2

Query Match 7.8%; Score 64.5; DB 4; Length 83;
Best Local Similarity 29.3%; Pred. No. 1.3;
Matches 22; Conservative 10; Mismatches 24; Indels 19; Gaps 4;

OY 17 LLATLVQAVSAGAEFSSACRELGFSNLLCSCDILLGQFNLLQDLPDRCGCCQEEA 76
Db 11 LVATLVFAVYGQKEI---CKEL--TKPVKCS-----DPLCKLCKMEKER 53
OY 77 FETKLYAGALIEVC 91
Db 54 YEDGHCF--TILSKC 66

RESULT 11
US-09-352-990-27
; Sequence 27, Application US/09352990
; Patent No. 6253090
; GENERAL INFORMATION:
; APPLICANT: Farnodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase

;; FILE REFERENCE: BB-1191
;; CURRENT APPLICATION NUMBER: US/09/352,990
;; CURRENT FILING DATE: 1999-07-14
;; EARLIER APPLICATION NUMBER: 60/092,866
;; EARLIER FILING DATE: July 15, 1998
;; NUMBER OF SEQ ID NOS: 29
;; SOFTWARE: Microsoft Office 97
;; SEQ ID NO 27
;; LENGTH: 461
;; TYPE: PRF
;; ORGANISM: Escherichia coli
US-09-352-990-27

Query Match 7.8%; Score 64.5; DB 4; Length 461;
Best Local Similarity 29.3%; Pred. No. 15;
Matches 22; Conservative 13; Mismatches 17; Indels 23; Gaps 5;

OY 78 ETKKLYAGAI-LEVCXKL-----GR---FPVQAFVBSDKPKLFRGLQIKYRG- 123
Db 14 EFKPIHAGEVGHVYCGITVYDLCHIGHGRTFVAFDVARYLR-----FLGYKLKVRNI 67
OY 124 ---SDPYLKLDDNG 135
Db 68 TDDIKIIRANENG 82

RESULT 12
US-08-162-146-3
; Sequence 3, Application US/08162146
; Patent No. 5965788
; GENERAL INFORMATION:
; APPLICANT: HOUEBERNE, Louis-Marie
; APPLICANT: DEVINOY, Eve
; APPLICANT: THEPOT, Dominique
; TITLE OF INVENTION: Production of a Protein of Interest in
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,146
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00533
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/07179
; FILING DATE: 12-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELER: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-162-146-3

Query Match 7.6%; Score 63.5; DB 2; Length 127;
Best Local Similarity 26.7%; Pred. No. 3.3;
Matches 39; Conservative 13; Mismatches 45; Indels 49; Gaps 9;

QY 8 CLVP-AFGKRLATVIAVSAFAEFSSEA---CRELGSSNLLGSCDLLGOFNLIQL 63
DB 3 CLISIALG---LLA--LEAALALAPKFIAPVQVWCPEPSSSEETLCLIS----- 45
QY 64 DPDCRG---CCQEEAQETKLVAGALLEVCGXKLGFRFPVOAVF----- 105
DB 46 DNDCLGTVCCPSAAGSCS-----TPIIYPTPKAGRCFVQAPMLSQLCEELSDCANDI 100
QY 106 -RSDKPKLFRGLQIKYVRSDDPYLK 129
DB 101 ECRGDKKCCFSCAMRYL---EPITLE 123

RESULT 13

US-09-314-127-3
; Sequence 3, Application US/09314127
; Patent No. 6268545
; GENERAL INFORMATION:
; APPLICANT: HODDERLINE, Louis-Marie
; APPLICANT: DEVINOT, Eve
; APPLICANT: THEPOT, Dominique
; TITLE OF INVENTION: Production of a Protein of Interest in
; TITLE OF INVENTION: the Milk of a Transgenic Mammalian
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/314,127
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/162,146
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/07179
; FILING DATE: 12-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-314-127-3

Query Match 7.6%; Score 63.5; DB 4; Length 127;
Best Local Similarity 26.7%; Pred. No. 3.3;
Matches 39; Conservative 13; Mismatches 45; Indels 49; Gaps 9;

QY 8 CLVP-AFGKRLATVIAVSAFAEFSSEA---CRELGSSNLLGSCDLLGOFNLIQL 63
DB 3 CLISIALG---LLA--LEAALALAPKFIAPVQVWCPEPSSSEETLCLIS----- 45
QY 64 DPDCRG---CCQEEAQETKLVAGALLEVCGXKLGFRFPVOAVF----- 105
DB 46 DNDCLGTVCCPSAAGSCS-----TPIIYPTPKAGRCFVQAPMLSQLCEELSDCANDI 100
QY 106 -RSDKPKLFRGLQIKYVRSDDPYLK 129
DB 101 ECRGDKKCCFSCAMRYL---EPITLE 123

RESULT 14

US-08-444-792-4
; Sequence 4, Application US/08444792
; Patent No. 5726037
; GENERAL INFORMATION:
; APPLICANT: Bodary, Sarah C.
; APPLICANT: Gorman, Cornelia M.
; APPLICANT: McLean, John W.
; APPLICANT: Napier, Mary A.
; TITLE OF INVENTION: METHOD FOR PREPARING WATER SOLUBLE
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 720 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,792
; FILING DATE: 19-May-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380227
; FILING DATE: 30-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/218878
; FILING DATE: 28-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/821337
; FILING DATE: 13-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/444490
; FILING DATE: 01-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/290224
; FILING DATE: 22-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: P0552PIC3D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: Amino Acid
; TOPOLOGY: linear
; US-08-444-792-4

Query Match 7.6%; Score 63.5; DB 1; Length 718;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 15:32:20 ; Search time 43.02 Seconds

(without alignments)
361.843 Million cell updates/sec

Title: US-09-676-718a-1

Perfect score: 831

Sequence: 1 MAAGPSGCLVPAFGKRLLLA.....ILKWNDSVEEFLSEKLERI 162

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	226.5	27.3	158	2	E86191
2	81	9.7	653	2	S11448
3	81	9.7	668	2	T46574
4	81	9.7	680	2	S06158
5	80.5	9.7	661	2	A25398
6	80	9.6	269	2	C87482
7	78	9.4	290	2	S21877
8	78	9.4	292	2	S21878
9	78	9.4	484	1	SYRZER
10	76.5	9.2	558	2	D69644
11	76.5	9.2	633	2	S52727
12	76.5	9.2	664	2	T06598
13	76	9.1	265	2	S77181
14	76	9.1	668	2	S21880
15	75.5	9.1	261	2	S14875
16	74.5	9.0	516	2	S06443
17	74.5	9.0	636	2	JM0666
18	74.5	9.0	667	2	S24782
19	74.5	9.0	667	2	D44261
20	74	8.9	275	2	F82761
21	74	8.9	663	2	S20877
22	74	8.9	663	2	T38155
23	74	8.9	979	1	JC2349
24	74	8.9	996	2	I48721
25	73	8.8	273	2	T47601
26	73	8.8	293	2	S21881
27	73	8.8	346	2	A12229
28	73	8.8	1120	2	H71664
29	72.5	8.7	228	2	S17349

ALIGNMENTS

30	72.5	8.7	667	2	S21879	dnak-type molecule
31	72.5	8.7	830	2	T18860	hypothetical prote
32	72.5	8.7	1192	2	T48499	receptor-like prot
33	72	8.7	467	2	J00966	dnak-type molecule
34	72	8.7	663	2	T04078	dnak-type molecule
35	72	8.7	777	2	C85065	dnak-type molecule
36	71	8.5	421	2	F84450	kinasin-like prote
37	71	8.5	668	2	S71171	hypothetical prote
38	71	8.5	691	2	T32748	conserved hypothet
39	71	8.5	212	2	H84979	hypothetical prote
40	70.5	8.5	272	2	S59389	hypothetical prote
41	70.5	8.5	432	2	S71074	probable membrane
42	70.5	8.5	432	2	S71074	asparagine--cRNA 1
43	70.5	8.5	432	2	S71072	asparagine--cRNA 1
44	70.5	8.5	644	2	S67431	dnak-type molecule
45	70.5	8.5	658	2	S38890	dnak-type molecule

RESULT 1

E86191

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: E86191

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Corne, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID: 21016719

A:Accession: E86191

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-158 <STO>

A:Cross-references: GB:AE005172; MID:94836919; PIDN:AA030621.1; GSPDB:GNO0141

C:Genetics:

A:Map position: 1

Query Match 27.3%; Score 226.5; DB 2; Length 158;

Best Local Similarity 32.9%; Pred. No. 6.1e-15;

Matches 48; Conservative 37; Mismatches 54; Indels 7; Gaps 4;

17 LLLATVIAQVAFGAEFFSEACREAGFGSSNLCCSDLLGOF-NILQIDPPRCGCCQERA 75

DB 13 LLLATVIAQVAFGAEFFSEACREAGFGSSNLCCSDLLGOF-NILQIDPPRCGCCQERA 68

QY 76 QPEYTKL-YAGAILEYCGAKLGFPOVOAFVRSNPKLFRGLQIKYVRSNPKLGLDDN 134

DB 69 EDSMSVITYSGAILVECMKLVFPYRIVGFIEEKEK-FPSYKVOYIYFSPKRLIMDED 127

QY 135 GNIAEELSLKWNDSVEEFLSEKLE 160

DB 128 GEKESIRIDNKRREHLQYREKVK 153

RESULT 2

dnak-type molecular chaperone hsc70 - leishmania donovani

N:Alternate names: heat shock cognate protein 70

C:Species: Leishmania donovani

C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999

C:Accession: S11448; S11572

R:MacFarlane, J.; Blaxter, M.L.; Bishop, R.P.; Miles, M.A.; Kelly, J.M.

Eur. J. Biochem. 190, 377-384, 1990

A:Title: Identification and characterisation of a Leishmania donovani antigen belonging to the 39 kDa family of proteins
A:Reference number: S11448; MUID:90306037
A:Accession: S11448
A:Molecule type: DNA
A:Residues: 1-653 <MAC1>
A:Cross-references: EMBL:X52314
R:MacFarlane, J.
submitted to the EMBL Data Library, April 1990
A:Reference number: S11572
A:Accession: S11572
A:Molecule type: DNA
A:Residues: 1'208,'S',210-653 <MAC2>
A:Cross-references: EMBL:X52314; NID:g9495; PIDN:CAA36551.1; PID:g9496
C:Genetics:
A:Gene: hsc70
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 9.7%; Score 81; DB 2, Length 653;
 Best Local Similarity 23.4%; Pred. No. 6.3;
 Matches 36; Conservative 23; Mismatches 53; Indels 42; Gaps

```

QY 48 LCSSCDLIGENLIQLDPDRCGCCQOEAFPTKRLVGAILEVCXKILGRPPQVQAVRS 107
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 451 MTKDCHLGTDLGSLIPAPRGVQLEVTFD--LDANGILNVAEKEGKGKRRQIITIN 507
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 108 DKPKLIFR-----GLQ-----IKYRGSDPVLKLDIN 134
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 DKGRLSKDEIRMNVDAMKRYEADDRACDRVYEAKNGLLENAYSMKKNTLGSDSNVSGKLDS 567
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 135 --GNIAEELSI-LKW---NTBSEVEPLSEKLERI 162
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 568 DKATLNKEIDVYTLWELSSNQATREYEHKKEL 601
      | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  
```

```

RESULT 3
T46574
dnak-type molecular chaperone B1P precursor [similarity] - soybean
N/Alternate names: endoplasmic reticulum HSC70-cognate binding protein
C/Species: Glycine max (soybean)
C/Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 11-May-2000
C/Accession: T46574
B/Figueiredo, J.E.F.; Cascardo, J.M.; Carolino, S.M.B.; Alvin, F.; Fontes, E.P.B.
Braz. J. Plant Physiol. 9, 103-110, 1997
A/Title: Water-stress regulation and molecular analysis of the soybean B1P gene family
A/Reference number: Z23087
A/Accession: T46574
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-668 <FIG>
A/Cross-references: EMBL:AF031241; PIDN:AA86942.1
A/Experimental source: strain Rooncke; clone cUFVb1P1; seed
C/Genetics:
A:Gene: B1P
C:Superfamily: heat shock protein 70
C/Keywords: ATP; endoplasmic reticulum; molecular chaperone; stress-induced protein
F:1-27/homann: signal sequence #status predicted <SIS>
F:28-668/product: dnak-type molecular chaperone B1P #status predicted <MAT>

```

```

Query Match          9.7%  Score 81:  DB 2:  Length 666:
Best Local Similarity 22.5%:  Pred. No. 6.5;
Matches 40;  Conservative 26;  Mismatches 63;  Indels 50;  Gaps 7.
OY  5  PSCGLVPAFGKRLLATVLIQ-----AVSAFGAEFSSSEACRELGSSNLCSSCDLLGQF  58
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db  446 PNTVIPT--KKSQVFTTQDOQTIVSIQVFEGRS-----LTKDCRLTKGRF  490
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  59 NLTQDPPCRGCCGHEAQETFKKLAVGATLLECGKKLGRFPQVQAFVRSBDRKLPFRGLQI  118
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

[illegible]

RESULT 4

506158 dnak-type molecular chaperone hsp70 - Trypanosoma cruzi

N:Alternate names: heat shock protein 70

C:Species: Trypanosoma cruzi

C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 15-Oct-1999

C:Accession: S06158; A28531, S25649; A44979

R:Bedate, C.A.

submitted to the EMBL Data Library, December 1988

A:Reference number: S06158

A:Accession: S06158

A:Molecule type: DNA

A:Residues: 1-680 <BRD>

A:Cross-references: GB:X07083; EMBL:X13690; NID:g10621; PIDN:CAA30115.1; PID:g10622

A:Experimental source: strain Maracal

R:Queuena, J.M.; Lopez, M.C.; Jimenez-Ruiz, A.; de la Torre, J.C.; Alonso, C.

Nucleic Acids Res. 16, 1393-1406, 1988

A:Title: A head-to-tail tandem organization of hsp70 genes in Trypanosoma cruzi.

A:Reference number: A28531; MUID:88157703

A:Accession: A28531

A:Molecule type: DNA

A:Residues: 222-473, 'L', 475-521, 550-622, 'A', 624-630, 'S', 632-680 <REQ>

A:Cross-references: GB:X07083; GB:X13690; NID:g10621; PIDN:CAA30115.1; PID:g10622

A:Experimental source: strain Maracal

A:Note: The sequence is revised in GenBank entry TCHSP70, release 111.0, (PIDN:CAA301 submitted to the EMBL Data Library, August 1992

A:Description: Cloning and characterization of cDNA encoding glucose-regulated protein

A:Reference number: S25648

A:Accession: S25649

A:Molecule type: mRNA

A:Residues: 1-46, 'T', 48-63, 'T', 65, 'Sp', 68-167, 'M', 169-178, 'R', 180-240, 'S', 242-256, 'T', 518-522, 'S', 524-529, 'SQ', 533-533, 'E', 535, 'RRR', 539-551, 'T', 556-557, 'E', 559-570, 'N', 57

A:Cross-references: EMBL:X67716; NID:g10625; PIDN:CAA47952.1; PID:g10626

A:Experimental source: strain Brazil

R:Engman, D.M.; Sias, S.R.; Gabe, J.D.; Donelson, J.E.; Dragon, E.A.

Mol. Biochem. Parasitol. 37, 285-287, 1989

A:Title: Comparison of HSP70 genes from two strains of Trypanosoma cruzi.

A:Reference number: A44979; MUID:90114336

A:Accession: A44979

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-46, 'T', 48-63, 'T', 65-167, 'M', 169-240, 'S', 242-256, 'T', 258, 'SQ', 261-310, 'D', 'SQ', 532-533, 'E', 535, 'RRR', 539-551, 'V', 553-554, 'T', 556-557, 'E', 559-570, 'N', 572-581, 'Q

A:Cross-references: GB:M6355; NID:g162116; PIDN:AAA30205.1; PID:g162117

A:Experimental source: strain Peru

C:Genetics:

A:Gene: hsp70

C:Function:

A:Description: involved in protein folding and assembling/disassembling of protein co

C:Superfamily: heat shock protein 70

C:Keywords: ATP; molecular chaperone

Query Match 9.7%; Score 81; DB 2; Length 680;

Best Local Similarity 27.8%; Pred. No. 6.6;

Matches 37; Conservative 16; Mismatches 52; Indels 28; Gaps 4;

DB 48 LCSSCDLIGONFLLDLPDCCRCCEAEAFETKKIYAGAILVECGXKIGRPQVQAFVRS 107

432 MKKCHLLGTGTELSIRPPRGVPIETVTPD--LDANGILNVAEGEGCTKRNDYIVLTN 508

DB 108 DKPKIFRGLQIKYVR-----GSDPYLKLLDNG-----NIAEEL--S 142

509 DKGRISRAEIRMEVBEAAKYEAEDDOVROLDANKNGLENTAFSKMNAVNDVNGAKIEEA 568

QY 143 ILKMTDSEVEEL 155
DB 569 DKRTTSAVEAL 581

RESULT 5

A25398
dnak-type molecular chaperone - Trypanosoma brucei
N:Alternate names: heat shock protein 70
C:Species: Trypanosoma brucei
C>Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 13-Mar-1998
C:Accession: A25398
R:Glass, D.J.; Polvere, R.I.; Van der Ploeg, L.H.T.
Mol. Cell. Biol. 6, 4657-4666, 1986
A:title: Conserved sequences and transcription of the hsp70 gene family in Trypanosoma
A:Reference number: A25398; MUID:87089827
A:Accession: A25398
A:Molecule type: DNA
A:Residues: 1-661 <GLA>
A:Cross-references: GB:M14697
C:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 9.7%; Score 80.5; DB 2; Length 661;
Best Local Similarity 24.7%; Pred. No. 7.2;

Matches 36; Conservative 16; Mismatches 55; Indels 39; Gaps 4;

QY 48 LCSSCDLQGFNLQDPPCRGCCQEAEQETKLYAGAILVEYCGKLGKRPQVQAFVRS 107
DB 452 MKRDCILGTFDLISGIPAPRGVQIEVTFD---LDANGILSYSAEKGKGNQIYITN 508
QY 108 DRPKLFR-----GLQ-----IKYVRGSDPYLKLDDN 134
DB 509 DKGRLSKAIERWVSAAYTEADKAHVXXIDAKNGLENYAFSMKNTINDPNVAGKLDDA 568
QY 135 GNIAEELSI--LKMTDSEVEELSE 157
DB 569 DKNAVTVTAVEALRWLNDQNEASLEE 594

RESULT 6

C87482
this protein [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: C87482
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87482
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:AE005673; NID:g1342327; PIDN:AAK23855.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCI880
C:Superfamily: thiamin biosynthesis protein thig

Query Match 9.6%; Score 80; DB 2; Length 269;
Best Local Similarity 29.4%; Pred. No. 2.9;

Matches 42; Conservative 15; Mismatches 66; Indels 20; Gaps 5;

QY 9 LVPATGKRLLATVLAQVAFGAFFSSACRELGFSNNLLCSCDLL-----GQFNLLQ 62
DB 29 LIVGCKRYDYKATNMAAAAGAEIYTVAVRRV-----NLTPSQPLLVIVYKTFEFTYL- 83

QY 63 LDPDCRCCQEOEAQFETKRLYAGA-----ILEYCGKLGKRPQVQAFVRSNDRKLPRL 116
DB 84 --PNTAGCFTGEDAVRTLRAREAGMDLVKLEVLSDPKTLTFPDMETLSIKLLVADGF 141

QY 117 QIKYVRGSDPY-LKLDDNGNIA 138
DB 142 QVMVYCSDDPYVARKLEEAGAVA 164

RESULT 7

S21877
dnak-type molecular chaperone b1p1 - common tobacco (fragment)
N:Alternate names: luminal binding protein b1p1
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
C:Accession: S21877; PQ0264
R:Denecke, J.; Goldman, M.H.; Demolder, J.; Seurinck, J.; Botterman, J.
submitted to the EMBL Data Library, June 1991
A:Description: The luminal binding protein (b1p) is encoded by a multigene family in
A:Reference number: S21877
A:Accession: S21877
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-290 <DEN>
A:Cross-references: EMBL:X60060; NID:g19804; PIDN:CAA42662.1; PID:g19805
R:Denecke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
Plant Cell 3, 1025-1035, 1991
A:title: The tobacco luminal binding protein is encoded by a multigene family.
A:Reference number: JQ1360; MUID:92361242
A:Accession: PQ0264
A:Molecule type: mRNA
A:Residues: 263-290 <DEW>
A>Note: translation of the nucleotide sequence is not complete
C:Genetics:
A:Gene: b1p1
A:Function:
A:Description: Involved in protein folding and assembling/disassembling of protein co
uration steps in vivo
C:Superfamily: heat shock protein 70
C:Keywords: ATP; endoplasmic reticulum; molecular chaperone
F;287-290/Region: endoplasmic reticulum retention signal

Query Match 9.4%; Score 78; DB 2; Length 290;
Best Local Similarity 23.0%; Pred. No. 5;
Matches 41; Conservative 24; Mismatches 63; Indels 50; Gaps 7;

QY 5 PSCCLVPACFKRLLLATVLA-----AVSAFGAFSSSEACRELGFSNNLLCSCDLLGQF 58
DB 71 PRNTVPIPT--KKSQVFTTYDQDOTVTSIQVEGERS-----LTKDCRLLGKF 115
QY 59 NLQDPPCRGCCQEOEAQFETKRLYAGAILVEYCGKLGKRPQVQAFVRSRKPRLFRQI 118
DB 116 DLGVIAPAPRGVQIEVTFD--ANGILNVAEDKGTGKSKRTITNDKGR----- 166
QY 119 KYVRGSDPYLKLDDNGNIAEELSIKMTND---SVEEF-----ISEKLE 160
DB 167 ----SQEIERMVRLEAEFEEDKRYKERIDANSLETTYNNKKNQINDKRLADKLE 220

RESULT 8

S21878
dnak-type molecular chaperone b1p2 - common tobacco (fragment)
N:Alternate names: luminal binding protein b1p2
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
C:Accession: PQ0262; S21878
R:Denecke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
Plant Cell 3, 1025-1035, 1991
A:title: The tobacco luminal binding protein is encoded by a multigene family.
A:Reference number: JQ1360; MUID:92361242
A:Accession: PQ0262
A:Molecule type: mRNA

A.: Ehlich, S.D.; Emmerson, P.T.; Ertian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal
lechi, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Huilo,
Koetter, P.; Koningsfeld, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, X.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scifone, P.; Seigunchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsita, P.; Toyonoh, A.; Tosato, V.; Uchuya
T.; Winters, P.; Wiipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033

A:Accession: D69644

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-558 <KUN>

A:Cross-references: GB:J09115; GB:AL009126; NID:g2634478; PIDN:CAB14105.1; PID:g26346

A:Experimental source: strain 168

C:Genetics:

A:Gene: llyd

C:Superfamily: dihydroxy-acid dehydratase

Query Match 9.2%; Score 76.5; DB 2; Length 558;
Best Local Similarity 21.5%; Pred. No. 15;
Matches 37; Conservative 33; Mismatches 75; Indels 27; Gaps 7;

OY 1 MAAPSGCLVAFPCFKRLTLATVLQAVSAFGA-EFSSSEACRELGFPSSNILSCDILGOFN 59
Db 143 VSQGPMAAGRISTYRKISLSVEFGVGAYQAKINENDELDEPGCPCTGCSGS--CHFT 200
OY 60 LLQDPDCRGCCOEAEQETKKLYAGAILLEVCGXKLGRFPVOQAIVRSDBKPLFRGLQIK 119
Db 201 ANSNH-----CLSEA-----LGLALPONGITLATSPERKEFVKSAQAQLMETIR-K 245
OY 120 YVRSDPY-LKLDD-----NGNTAEELSTLKMTWDSVEEFLSEKLERI 162
Db 246 DIKPRDIYTVKAINDMAPALDMALGSGTNTVLTALTALANEAEGVEYSLEIRINEV 297

RESULT 11
S52727

dnaK-type molecular chaperone hsp70 - Leishmania donovani infantum (fragment)
N:Alternate names: heat shock protein hsp70; immunodominant antigen
C:Species: Leishmania donovani infantum
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 20-Aug-1999
C:Accession: S52727
R:Quijada, L.; Requena, J.M.; Soto, M.; Alonso, C.
Submitted to the EMBL Data Library, March 1995
A:Description: In canine viscero-cutaneous leishmaniasis the anti-hsp70 antibodies ar
A:Reference number: S52727
A:Accession: S52727
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-653 <DU1>
A:Cross-references: EMBL:X85798; NID:g758135; PIDN:CAA59793.1; PID:g758136
C:Genetics:
A:Gene: hsp70
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein co
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 9.2%; Score 76.5; DB 2; Length 653;
Best Local Similarity 27.9%; Pred. No. 18;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDILGONILQDPDCRGCCOEAEQETKKLYAGAILLEVCGXKLGRFPVOQAIVRS 107
Db 451 MKDKCHLGTGDTSLGIPAPRGLPQIEVTPT---LDAGILLVAEEGGTKRMQDITTN 507

A.: Ehlich, S.D.; Emmerson, P.T.; Ertian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gal
lechi, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Huilo,
Koetter, P.; Koningsfeld, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, X.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scrofione, P.; Serighini, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsita, P.; Toyonoh, A.; Tosato, V.; Uchuya
T.; Winters, P.; Wiipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zanchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033

A:Accession: D69644

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-558 <KUN>

A:Cross-references: GB:J09115; GB:AL009126; NID:g2634478; PIDN:CAB14105.1; PID:g26346

A:Experimental source: strain 168

C:Genetics:

A:Gene: llyd

C:Superfamily: dihydroxy-acid dehydratase

Query Match 9.2%; Score 76.5; DB 2; Length 558;
Best Local Similarity 21.5%; Pred. No. 15;
Matches 37; Conservative 33; Mismatches 75; Indels 27; Gaps 7;

OY 1 MAAPSGCLVAFPCFKRLTLATVLQAVSAFGA-EFSSSEACRELGFSSNILSCDILGOFN 59
Db 143 VSQGPMAAGRISTYRKISLSVEFGVGAYQAKINENDELDEFGCPCTGCSCS--CHFT 200
OY 60 LLQDPDCRGCCOEAEQETKKLYAGAILLEVCGXKLGRFPVOQAIVRSDBKPLFRGLQIK 119
Db 201 ANSNH-----CLSEA-----LGLALPONGITLATSPERKEFVKSAQAQLMETIR-K 245
OY 120 YVRSDPY-LKLDD-----NGNTAEELSTLKWMTDSVEEFLSEKLERI 162
Db 246 DIKPRDIYTVKAINDMAPALDMALGSGTNTVLTALTALANEAQGVEXSLERINEV 297

RESULT 11
S52727

dnaf-type molecular chaperone hsp70 - leishmania donovani infantum (fragment)
N:Alternate names: heat shock protein hsp70; immunodominant antigen
C:Species: Leishmania donovani infantum
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 20-Aug-1999
C:Accession: S52727
R:Quijada, L.; Requena, J.M.; Soto, M.; Alonso, C.
Submitted to the EMBL Data Library, March 1995
A:Description: In canine viscero-cutaneous leishmaniasis the anti-hsp70 antibodies ar
A:Reference number: S52727
A:Accession: S52727
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-653 <DUi>
A:Cross-references: EMBL:X85798; NID:g758135; PIDN:CAA59793.1; PID:g758136
C:Genetics:
A:Gene: hsp70
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein co
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone

Query Match 9.2%; Score 76.5; DB 2; Length 653;
Best Local Similarity 27.9%; Pred. No. 18;
Matches 24; Conservative 14; Mismatches 35; Indels 13; Gaps 2;

OY 48 LCSSCDILGONILQDPDCRGCCOEAEQETKKLYAGAILLEVCGXKLGRFPVOQAIVRS 107
Db 451 MKDKCHLGITDLSICIPARPGLPQIEVTPT---LDAGILLVAEEEGTKRQMIIITTN 507

QY 108 DKPKLFRGLQIRYKRGSDPVKLDD 133
 Db 508 DKGRLL-----SKDEIERMVD 523

RESULT 12

T06598
 dnak-type molecular chaperone Bip-A - soybean
 N:Alternate names: binding protein Bip isoform A
 C:Species: Glycine max (soybean)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: T06598
 R:Kalinowski, A.; Kowley, D.L.; Loer, D.S.; Foley, C.; Buta, G.; Herman, E.M.
 A:Title: Binding-protein expression is subject to temporal, developmental and stress-ind
 A:Reference number: 215624; MUID:95218610
 A:Accession: T06598
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-664 <KAL>
 A:Cross-references: EMBL:U08384; NID:q475601; PIDN:AAA81956.1; PID:q475602
 A:Experimental source: strain Century 84; leaf
 C:Function:
 A:Description: involved in protein folding and assembling/disassembling of protein comp
 C:Superfamily: heat shock protein 70
 C:Keywords: ATP; molecular chaperone

Query Match 9.2%; Score 76.5; DB 2; Length 664;
 Best Local Similarity 22.2%; Pred. No. 18;

Matches 45; Conservative 26; Mismatches 69; Indels 63; Gaps 8;

QY 5 PSCGLVPAGKRLATVLYQ-----AVSAGFAEFSSEACRELGFSSNLLCSSCDLLGQF 58
 Db 444 PRNTVIPT--KKSQVFTTYQDOQTIVTSVFEGERS-----LTKDCRLGKF 488
 QY 59 NLLQDPPRCGCCOEAPETKRLVAGALVEYCGKILGFPOVOAFVRSDKRLFRGLQI 118
 Db 489 ELSGIPAPRGTPQIEVFEVD--ANGILNKAEDKSGKSEKITITNEKRLSQEIE 545
 QY 119 KYVRGSD-----PVLKLD-----DNGNIAEL----- 141
 Db 546 RMVRAEAEFAEDKKVKRIRIDARNSLETYYVMKNQVSDKDLKLESDEKEKVEYAVK 605
 QY 142 SILKMTD--SVEEFLSEKLERI 162
 Db 606 EALEWLDNQGSEYKRYEKKKEV 628

RESULT 13

S77181
 hypothetical protein slr1809 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S77181
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-116, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S77181
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-265 <KAN>
 A:Cross-references: EMBL:D090908; GB:AA001339; NID:q1652725; PIDN:BA117739.1; PID:0101847
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 9.1%; Score 76; DB 2; Length 265;
 Best Local Similarity 23.2%; Pred. No. 7.1;

Matches 41; Conservative 22; Mismatches 60; Indels 54; Gaps 8;

QY 31 AEFSSACRELGFSSNLLCSSCDLLGQF-----NLLQDPPRCGCCOE 73
 Db 74 AELDQIAFNPNPESKRLSSAEAYLGKMKQHRRAPDQMHRLVRGCLSLPECCBQALD 133
 QY 74 EAQFETK-----LVGALVEYCGKIL-----GRPOVOAFVRSDKRLFRGLQ 117
 Db 134 LAEFRRQRKRYRTYVSGILMFAAIL--CGGGIYLLSQGLP---FVIEKRL-ENIQ 186
 QY 118 IYVRGSDPVKLDDNNGNIAELSTL-----KMTDSVEEF-----LSERLE 160
 Db 187 STTEAMNDITIAQLMGQEALEALREPTLLISQQLQENLNDRLENLQARLKALEEKIK 243

RESULT 14

S21880
 dnak-type molecular chaperone b1p5 precursor - common tobacco
 N:Alternate names: luminal binding protein b1p5
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
 C:Accession: S21880; J01361
 R:Denoeke, J.; Goldman, M.H.; Demolder, J.; Seurinck, J.; Botterman, J.
 Submitted to the EMBL Data Library, June 1991
 A:Description: The luminal binding protein (b1p) is encoded by a multigene family in
 A:Reference number: S21877
 A:Accession: S21880
 A:Molecule type: mRNA
 A:Residues: 1-668 <DEN>
 A:Cross-references: EMBL:X60058; NID:q19812; PIDN:CAA42660.1; PID:q19813
 R:Denoeke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.
 Plant Cell 3, 1025-1035, 1991
 A:Title: The tobacco luminal binding protein is encoded by a multigene family.
 A:Reference number: J01360; MUID:92361242
 A:Accession: J01361
 A:Molecule type: mRNA
 A:Residues: 1-653, 'G', 655-668 <DEN>
 A:Note: translation of the nucleotide sequence is not complete
 C:Genetics:
 A:Gene: b1p5
 C:Function:
 A:Description: involved in protein folding and assembling/disassembling of protein co
 uration steps in vivo
 C:Superfamily: heat shock protein 70
 C:Keywords: ATP; endoplasmic reticulum; molecular chaperone
 F.1-24/Domain: signal sequence #status predicted <SIG>
 F.665-668/Region: endoplasmic reticulum retention signal

Query Match 9.1%; Score 76; DB 2; Length 668;
 Best Local Similarity 23.0%; Pred. No. 20;

Matches 41; Conservative 23; Mismatches 64; Indels 50; Gaps 7;

QY 5 PSCGLVPAGKRLATVLYQ-----AVSAGFAEFSSEACRELGFSSNLLCSSCDLLGQF 58
 Db 447 PRNTVIPT--KKSQVFTTYQDOQTIVTSVFEGERS-----LTKDCRLGKF 491
 QY 59 NLLQDPPRCGCCOEAPETKRLVAGALVEYCGKILGFPOVOAFVRSDKRLFRGLQI 118
 Db 492 DLTGTPAPRGTPQIEVFEVD--ANGILNKAEDKSGKSEKITITINDKRL----- 542
 QY 119 KYVRGSDPVKLDDNNGNIAELSTLKMNTD--SVEEF-----LSERLE 160
 Db 543 -----SQEIERMVAKEAEFAEDKKVKRIRIDARNSLETYYVMKNQVSDKDLKLESDEKEKVEYAVK 596

RESULT 15

S14875
 dnak-type molecular chaperone hsp70 - Trypanosoma cruzi (fragment)
 N:Alternate names: heat shock protein 70
 C:Species: Trypanosoma cruzi
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 20-Aug-1999
 C:Accession: S14875

R; Levy Yeyati, P.; Lafon, S.; Bonnefoy, S.; Levi, M.
submitted to the EMRI Data Library April 1991

submitted to the EMBL Data Library, April 1991

A; Description: Nucleotide sequence of a Trypanosoma cruzi HSP-70 c-DNA.

A; Reference number: S14875

A;Accession: S14875

A; molecule type: mRNA

A;Residues: 1-261 <LEV>

A; Cross-references: EMBL:X58715; NID:g10623; PIDN:CAA41551.1; PID:g10624

C;Genetics:

A; Gene: hsp

C;Function:

A;Description: Involved in protein f

C; Superfamily: heat shock protein 70

Query Match	9.18;	Score 75.5;	DB 2;	Length 261;
Best local similarity	33.88;	Need No	7.8;	

Best Local Similarity 33.8%; Pred. No. 7.8;

Matches 22; Conservative 9; Mismatches 31; Indels 3; Gaps 1;

QY 48 LCSSCDLGFNLLQLDPDCRGCCQEEAQETKKLYAGAILLEVCGKKLGRFPQVQAFVRS 107

Db 52 MTKDCHLTGFDLSGIPAPRGVPQIEVTD--LDANGILNVSAAEEKGTGKRNQIVITN 108

QY 108 DKPKL 112

Db 109 DKGRL 113

Search completed: September 18, 2002, 16:33:51

Job time: 3691 sec